

```

Db      1654 GGGCTGCTGCTCCAGAGATTTCAGAGTCTCTAACAAGAGAGATGCAATTACCAAGT 1713
Qy      501 G1ygluYrprGluValleuG1yArgh1sG1ySerAn11eg1nva1AerG1uValArgrArg 520
Db      1714 GGTGAATGGAAAGTTCTTGCGCGGCGATCATATATCCAAAGTTGAAGAGCAAGG 1773
Qy      521 LeuValTyrPheG1uG1YrThrLyAspSerProleuG1uH1sH1sleuTyrValValSer 540
Db      1774 CTGGATATATTTTGAAGGACCAAGACATCCCTTTAGGACATCACTGTAGCTAGTCAGT 1833
Qy      541 TyrValAsnProG1yG1uValThrArgleuThrAspArg1yTyrSerH1sSerCysCys 560
Db      1834 TACGTAAATCCTGGAGAGAGTACAAAGCTGACCTGCGCTACTCACTTCTTGCTGC 1893
Qy      561 I1eserG1nH1sCyAspPhePhe11eser1ySTyrSerAenG1uLyAsnProH1sCys 580
Db      1894 ATCAGTCAGCAGCTGAGCTCTTATTAAGTAAGTACTAAACAGAAATCCACACAGT 1953
Qy      581 ValSerLeuTyrLyAspSerSerProG1uAspApproThrCysLyAspThrLyGluPhe 600
Db      1954 GTGTCCTTTACAGAGCTATCAAGTCTCAAGATGACCCCACTTGCAAAACAAAGGAAATT 2013
Qy      601 TrpAlaThr11eLeuAspSerAlaG1yProleuProAspTyrThrProProG1u11ePhe 620
Db      2014 TGGGCGACCATTTTGGATTTCAGCAGGTCCTCTCTGACTATACCTCCCAAGAAATTTTC 2073
Qy      621 SerPheG1uSerThrThrG1yPheThr1eutyG1yMetLeuTyrLyAspProH1sAspLeu 640
Db      2074 TCTTTTGAAGTACTACTGAGATTATCATTTGATGGAGTGCCTCTCAACCTCATGATCTTA 2133
Qy      641 G1nProG1yLyAspTyrProThrValLeuPhe11eTyrG1yG1yProG1uValG1uLeu 660
Db      2134 CAGCGTGAAGAAATATCTCTACTGTGCTGTTCATATATGGTGTCTCTCAGGTGAGTGTG 2193
Qy      661 ValAsnAsnArgPheLySG1yValLySTyrPheArgLeuAenThrLeuAlaSerLeuG1y 680
Db      2194 GTGATATATCGGTTTAAAGAGAGTCAAGTATTTCCGCTTGAAATCCCTAGCCTCTAGCT 2253
Qy      681 TyrValValValVal11eAspAsnArgG1ySerCyH1sAsArg1yLeuLyPheG1uG1y 700
Db      2254 TATGTGCTGTGATGATGAGAAACAGGGGATCTCTGTCCAGAGGGCTTTAAATTTGAAAGGC 2313
Qy      701 AlaPheLySTyrLyMetG1yG1n11eG1u11eAspAspArg1nValG1uG1yLeuG1nTyr 720
Db      2314 GCCTTTAAATATTAATGGGTCAAATTAAGATGACGTACGGTGGAGAGACTCCAAATAT 2373
Qy      721 LeuAlaSerArgTyrAspPhe11eAspLeuAspArgValG1y11eH1sG1yTrpSerTyr 740
Db      2374 CTACCTTCTGATATGATTTTATTGACTTAGTGTGTGGGACATCCAGCGTGGTCCAT 2433
Qy      741 G1yG1yTyrLeuSerLeuMetAlaLeuMetG1nArgSerArgP11ePheArgValAla11e 760
Db      2434 GGAGAGATACCTCTCCCGATGGCATTAATGACAGAGTCAAGATCTTCAAGGTTGCTAAT 2493
Qy      761 AlaG1yAlaProValThrLeuTrp11ePheTyrAspThrG1yTyrThG1uArgTyrMet 780
Db      2494 GCTGGGGCCCAAGTACCTCTGATCTTCTATATGATACAGATACAGGAACGTTATATG 2553
Qy      781 G1yH1sProAspG1nAsnG1uG1nG1yTyrTyrLeuG1ySerValAlaMetG1nAlaG1u 800
Db      2554 GGTACACCTGACCAAGATGAACAGGGCTATTACTTAGATCTGTGGCCATCAAGACAA 2613
Qy      801 LysPheProSerG1uProAsnArgLeuLeuLeuH1sG1yPheLeuAspG1uAsnVal 820
Db      2614 AAGTTCCTCTGAAACCAATGTTTACTGCTTACATGGTTCCTGGAAGAAGATATC 2673
Qy      821 HisPheAlaH1sThrSer11eLeuLeuSerPheLeuValArgAlaG1yLyAspProTyrAsp 840
Db      2674 CATTTTGACADACAGATATATTACTGAGTTTTTAGAGAGGGCTGGAACCATATGAT 2733
Qy      841 LeuG1n11eTyrProG1nG1uArgH1sSer11eArgVal1ProG1uSerG1yG1uH1sTyr 860

```

```

Db      2734 TTACAGATCTATCTCTCAGAGAGACACAGCATAAAGATTCCTGAATCGGAGACATTAT 2793
Qy      861 G1uLeuH1sleuLeuH1sTyrTyrLeuG1nG1uAsnleuG1ySerArg11eAla11eLeuLyS 880
Db      2794 GAATCGATCTTTTGGACTACTACCTTCAAGAAACCTTGATACGTAATGCTGCTTAA 2853
Qy      881 Val11e 882
Db      2854 GTGATTA 2859

RESULT 6
AAH99934
ID AAH99934 standard; cDNA; 3143 BP.
AC AAH99934;
XX
XX
XX 12-APR-2002 (first entry)
DE cDNA encoding 21953 human prollyl oligopeptidase.
XX
XX 21953 prollyl oligopeptidase; human; proline; endopeptidase;
XX cancer; cardiovascular disease; autoimmune disease; atopic allergy;
XX neuronal disorder; vascular disorder; prostate disorder; cytostatic;
XX antidiabetic; antiarthritic; antiaesthetic; antiinflammatory;
XX diabetes mellitus; arthritis; multiple sclerosis; asthma;
XX Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
XX Homo sapiens.
XX 0
XX Key Location/Qualifiers
XX CDS 229..2877
XX FT /*tag= a
XX FT /product= "21953 prollyl oligopeptidase"
XX FT /note= "this region is specifically claimed in
XX FT Claim 2"
XX
XX MO200179473-A2.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001; 2001MO-US40483.
XX
XX 18-APR-2000; 2000US-197508P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Meyers RA, Williamson M;
XX
XX WPI; 2002-034353/04.
XX P-PSDB: AAG78415.
XX
XX New polypeptides 21953, member of human prollyl oligopeptidase family,
XX useful as diagnostic targets and therapeutic agents for controlling
XX cancer, lymphoma and leukemia
XX
XX
XX Claim 7; Page 100-102; 121pp; English.
XX
XX This invention relates to an isolated 21953 human prollyl
XX oligopeptidase. Which is cytosolic, antidiabetic, antiarthritic,
XX neuroprotective, antihypertoid, dermatological, antispasmodic,
XX antiaesthetic, ophthalmological, antiinflammatory, nootropic,
XX antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
XX antidiagonal, cardiant, antiatherosclerotic, anorectic and
XX metabolic in its action. Uses include gene therapy, expression or
XX activity of 21953 protein modulator, it is useful for identifying a
XX compound which binds to it and can be used in preventing, treating
XX or degrading a cellular proliferative or differentiative disorder.
XX The 21953 molecules can act as novel diagnostic targets and therapeutic
XX agents for controlling disorders associated with the aberrant activity
XX or degradation of peptide hormones e.g., disorders associated with cell
XX differentiation and proliferation such as cancer, immune function,
XX reproductive, neurological and cardiovascular function. The 21953
XX molecules are thus useful for treating and preventing cellular

```

CC proliferative and differentiative disorders, haematopoietic neoplastic  
CC disorders, immune disorders such as autoimmune diseases, diabetes  
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease, and  
CC neuronal disorders, demyelinating diseases, vascular disorders and  
CC metabolism or pain disorders. This sequence represents the cDNA  
CC encoding sequence of 21953 human prolyl oligopeptidase.

XX Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;

## Alignment Scores:

Pred. No.:	0	Length:	3143
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-070-464-1 (1-882) x AAH99934 (1-3143)

QY 1 MetAlaAlaIaIaMetGluThrGluGluIleuGluValGluIlePheGluThrAlaAspCys 20  
DB 229 ATGGCAGCAGCAATGGAAACAGAAACAGCTGGGTGTGAGATATTGAAACTGCCGACTGT 288  
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40  
DB 289 GAGGAGAAATATTGAATACAGGATCGGCTTAATTGAGCTTTTATGTGAGCGGAT 348  
QY 41 SerTrpSerGlnLeuLysLysLeuLeuAlaSerThrArgLysTyrHisGlyTyrMetMet 60  
DB 349 TCCGGAGTCAGCTTAAGAAAGCTCTTGCCATACCGAAATATCATGGCTACATATG 408  
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80  
DB 409 GCTAAGCACCAACATGATTTTCATGTTGTGAAGAGATATCCAGATGACCTCATTTCA 468  
QY 81 AspArgLleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100  
DB 469 GACGAAATCATTAACCTTGGCATGTCTGGTGAACAGAAATTAACACTGTTTATTTCT 528  
QY 101 GluIleProLysThrIleAsnArgAlaAlaLeuMetLeuSerTrpLysProLeuLeu 120  
DB 529 GAAATTCCTCAAACTTATCATAGAGACAGCTTAAATGCTCTTGGAAGCCCTTTTGG 588  
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140  
DB 589 GATCTTTTTCAGGCAACCTGGACTATGGAATGATTCGAGAGAAAGAACTATTAGA 648  
QY 141 GluArgLysArgIleGlyThrValGlyIleAsnSerTyrAspTyrHisGlnGlySerGly 160  
DB 649 GAAAGAAACGCAATGGAAACAGTCGGAATGCTTTCATCAGATTATCACCAAGGAAGTGA 708  
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180  
DB 709 ACATTTCTGTTTCAAGCCGGAGTGGAAATTAATCACTAAAGAGGAGGCCCAAGGA 768  
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
DB 769 TTATACGCAACACTTTAAGGCCCAATCTAGTGAACCTAGTTGCTCCACATATCGAATG 828  
QY 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220  
DB 829 GATCCAAATTAATGCTCTGCTGATCCAGACTGGAATGCTTTTATATACCAACAGATATT 888  
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240  
DB 889 TGGATATCTAACCTCGTACCAAGAGAAAGAGAGACTCTTATATGTCACCAATGAGCTA 948  
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260  
DB 949 GCCAATCGAAGAAAGATGCGCAGATCAGTGAAGTGGCTACTCTTTGTTCCAAAGAA 1008  
QY 261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280

DB 1009 TTGTATGATATTCGTGCTATTTGGTGGTGTCCAAAGCTGAACAACTCCAGTGTGCT 1068  
QY 281 LysIleLeuArgIleLeuTyrGluAsnAspGluSerGlyValGluIleIleHisVal 300  
DB 1069 AAAATTCCTTGAATTCATATGAAGAAATGATAACTGAGGTGAATTAATTAATGTT 1128  
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
DB 1129 ACATCCCTATGTTTGGAAACAGAGGAGGCAATTCCTTCCTTAAACAGTAC 1188  
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340  
DB 1189 GCAATTCCTAAAGCACTTTAAGATGTCAGAAATATATGATGATGTCGAAGGAAGATC 1248  
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360  
DB 1249 ATAGATGTCAATGATTAAGAACTAATTAACCTTTTATGAGATTCATTAATGAAGGATGAA 1308  
QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380  
DB 1309 TATATTCAGAGCTGATGATGACTCTCGAGGAAATATGCTTGTCATCTCATGAT 1368  
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
DB 1369 CGCTCCAGACTCGCTGCAATATGTTGATCTCACTGAATTAATTAATCCAGTAGAA 1428  
QY 401 AspAspValMetGluArgGluArgGluIleGluSerValProAspSerValThrProLeu 420  
DB 1429 GATGATGTATGAAAGCGAGACCTCATTAAGCACTGATGCTGATTTGTGAACCCACTA 1488  
QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440  
DB 1489 ATATATCTATGAAGAAACACACAGCATGTGATTAATTCATGACATCTTTCATGTTTT 1548  
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460  
DB 1549 CCCAAAGTCACGAAAGAGAAATGAGTTATTTTCTCTGAAATCAAAACAGGTTTC 1608  
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerTyrTyrLysArgSerSerGly 480  
DB 1609 CGTCATTTATTAACAAATTAATCATCTATTTTAAGGAAAGCAATTAACATCCAGTGT 1668  
QY 484 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500  
DB 1669 GGGCTGCTCTCCAAAGTATTTCAAGTCTCATTAACAAGAGAGATAGCAATTAACAGT 1728  
QY 501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520  
DB 1729 GTGGAATGGGAAGTTCTTGCCCGCATGATCTAATATCCAAATGATGAAGTCAGAAG 1788  
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSer 540  
DB 1789 CTGTATATTTTAAAGCACCAAGCACTCCCTTTAGAGATCACTGTATGATGATGAT 1848  
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560  
DB 1849 TACGTAAATCTGAGAGGAGGACAGAGCTGACCGTGCCTACTACATTTCTGCTGC 1908  
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580  
DB 1909 ATCAGTCAGACACTGTGACTTTTATATAGATATATATGTAACAGAAATCCACACTGT 1968  
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600  
DB 1969 GTGTCCCTTTAACAAGCTATCAAGTCTCGAAGATGAGCCCACTTGCAAAACAGAAATTT 2028  
QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620  
DB 2029 TGGGCAACATTTTGGATTCAGAGAGTCTCTTCTGACATATCTCTCCAGAAATTTTC 2088  
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640  
DB 2089 TCTTTGAAAGTACTACTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 2148

```

Y      641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
      2149 CACCTCGAAGAAATATCTACTGCTGCTGTCATATATGCTGCTCCTCAGGTCAGTTG 2208
Y      661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
      2209 GTGAATTAATCGGTTAAAGGAGTCAGATATTTCCGCTTAATACCTTAAGCTCTCAAGT 2268
Y      681 TyrValValValValIleLeuAsnArgGlySerCysHisArgGlyLeuLysPheGlyGly 700
      2269 TATGTGTTGTAAGTATGATACACACAGGAGATCCTGTCACGAGGCTTAATTTGAAGGC 2328
Y      701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyr 720
      2329 GCCTTTAAATTAATTAATGGTCAATATAGAAATGACATGACATGAGTGAAGGACCTCAATAT 2388
Y      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
      2389 CTAGCTTCGATATGATGATTTCAATGATTAATGATGATGATGATGATGATGATGATGATG 2448
Y      741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
      2449 GGAGGATACCTCTCCCTGATGATGATTAATGACAGAGTCAATATCTTCAGGCTGCTATAT 2508
Y      761 AlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlnArgTyrMet 780
      2509 GCTGGGGCCCCAGCTCCTGTCGATCTTCTATGATGATGATGATGATGATGATGATGATG 2568
Y      781 GlyHisProAspGlnAsnGlnGlnGlyTyrTyrTyrLeuGlySerValAlaMetGlnAlaGln 800
      2569 GGTCAACCTGACCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2628
Y      801 LysPheProSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnVal 820
      2629 AAGTTCCTCCCTGAACCAATCGTTTACTGCTCTTACATGATGATGATGATGATGATGATG 2688
Y      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
      2689 CATTTTCACATACCACTATATTAATGATGATGATGATGATGATGATGATGATGATGATG 2748
Y      841 LeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGlnSerGlyGlnHisTyr 860
      2749 TTACAGATCTATCTCTCAGAGACACAGCAATTAAGATTCCTGAATGGGAGAACATTAT 2808
Y      861 GlnLeuHisLeuLeuHisTyrLeuGlnGlnLeuAsnLeuGlySerArgIleAlaAlaLeuLys 880
      2809 GAACCTGCATCTTTGCACTACCTTCAAGAAACCTTGATCAGATGCTGCTCTAATAA 2868
Y      881 ValIle 882
      2869 GTGATTA 2874
Db

```

## RESULT 7

AAH99935 standard; cDNA; 2643 BP.

AAH99935:

12-APR-2002 (first entry)

Coding sequence of 21953 human prolyl oligopeptidase.

21953 prolyl oligopeptidase; antibody; proline; endopeptidase;  
 cancer; cardiovascular disease; autoimmune disease; atopic allergy;  
 neuronal disorder; vascular disorder; prostate disorder; cytostatic;  
 antidiabetic; anticholinergic; antidiabetic; antiinflammatory;  
 diabetes mellitus; arthritis; multiple sclerosis; asthma;  
 Grave's disease; neuronal disorder; demyelinating disease; ss.

XX Homo sapiens.  
 OS  
 XX  
 PN WO200179473-A2.

```

XX      25-OCT-2001.
PD      11-APR-2001; 2001WO-US40483.
XX      18-APR-2000; 2000US-197508P.
XX      (MILL-) MILLENNIUM PHARM INC.
XX      Meyers RA, Williamson M;
XX      WPI: 2002-034353/04.
XX      P-PSDB; AAG78415.
XX      New polypeptides 21953, member of human prolyl oligopeptidase family,
XX      useful as diagnostic targets and therapeutic agents for controlling
XX      cancer, lymphoma and leukemia
XX      Claim 2: Page: 121pp; English.
XX      This invention relates to an isolated 21953 human prolyl
XX      oligopeptidase, which is cytostatic, antidiabetic, antiarthritic,
XX      neuroprotective, antihypertensive, dermatological, antipsoriatic,
XX      antiasthmatic, ophthalmological, antiinflammatory, nootropic,
XX      antiparkinsonian, anticonvulsant, gynaecological, vasorelaxant,
XX      antidiabetic, cardiac, antithrombotic, anorectic and
XX      antitumor, in its action. Uses include gene therapy, expression or
XX      metabolic in its action. Uses include gene therapy, expression or
XX      compound which binds to it and can be used in preventing, treating
XX      or detecting a cellular proliferative or differentiative disorder.
XX      The 21953 molecules can act as novel diagnostic targets and therapeutic
XX      agents for controlling disorders associated with the aberrant activity
XX      or degradation of peptide hormones e.g., disorders associated with cell
XX      differentiation and proliferation such as cancer, immune function,
XX      reproductive, neurological and cardiovascular function. The 21953
XX      molecules are thus useful for treating and preventing cellular
XX      proliferative and differentiative disorders, haematopoietic neoplastic
XX      disorders, immune disorders such as autoimmune diseases, diabetes,
XX      mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
XX      neuronal disorders, demyelinating diseases, vascular disorders and
XX      metabolism or pain disorders. This sequence represents the cDNA
XX      encoding sequence of 21953 human prolyl oligopeptidase. This
XX      sequence represents the coding sequence of 21953, being the sequence
XX      in between the start and the stop codon of the sequence represented in
XX      AAH99934.
XX      Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;
XX      Alignment Scores:
XX      Pred. No.: 0 Length: 2643
XX      Score: 4695.00 Matches: 881
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 99.89% Indels: 0
XX      DB: 24 Gaps: 0
XX      US-10-070-464-1 (1-882) x AAH99935 (1-2643)
Y      2 AlaAlaIleMetGluThrGlnGlnLeuGlyValGlnIlePheGluThrAlaAspCysGlu 21
      1 GCGACGACGATGAGAAACAGAACAGCTGGGTGTTGAGATTAATGAAACATGCGGACGTGAG 60
Y      22 GlnAsnIleGlnSerGlnAspArgProLysLeuGlnGlnProPheTyrValGlnArgTyrSer 41
      61 GAGAAATATTAATGATCAAGATGCGCTTAATTAATGAGCCCTTTTATGTTGAGCGGATATTC 120
Y      42 TTPSerGlnLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMetAla 61
      121 TGGAGTCACCTTAATTAATGAGCTGCTGCGATACAGAAATATCAATGATGATGATGATG 180
Y      62 LysAlaProHisAspPheMetPheValLysArgAsnArgProAspGlyProHisSerAsp 81
      181 AAGGACACCATATGATTAATGTTTGTGAAGAGATGATGATGATGATGATGATGATGATG 240
Db

```

QY ArgIleTyrTyrLeuAlaMetSerGlyGluLeuArgGluLeuThrLeuPheTyrSerGlu 101  
 DB |||||  
 241 AGAATCTATACCTTGGCATGTCTGTGAGACAGAGAAATACCTGTTTATCTGAA 300  
 QY 102 IleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeuAsp 121  
 DB |||||  
 301 ATTCCCAADACATCAATAGAGACAGACAGCTTAAATGCTCTGTGAGACCTCTTTGGAT 360  
 QY 122 LeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArgGlu 141  
 DB |||||  
 361 CTTTTTCAGGCAACACTGACACTAGTAATCTCCAGAAAGAACTATTAAGACAA 420  
 QY 142 ArgLysArgGlyGlyTyrValGlyIleAlaSerTyrAspTyrHisGlySerGlyThr 161  
 DB |||||  
 421 AGAAAAACCATTTGGAACAGTCGGAATGCTCTTACGATATACACCAAGAGAGTGAACA 480  
 QY 162 PheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGlyPhe 181  
 DB |||||  
 481 TTTCTGTTCAAGCCGCTAGTGGATTTATACGTAATAAGATGAGGGCCACAGAGATT 540  
 QY 182 ThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMetAsp 201  
 DB |||||  
 541 ACGCAACCACTTTAAGGCCCAATCTAGTGAACCTAGTTGTTCCCAACATACGGATGGAT 600  
 QY 202 ProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIleTrp 221  
 DB |||||  
 601 CCAAAATTAATGCCCTGCTGATTCAGATCGAGATTGCTTTATACATAGCAAGCATATTTGG 660  
 QY 222 IleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeuAla 241  
 DB |||||  
 661 ATATCTAACATCGTAAACAGAGAAAGAGAGACTCATTTATGTGCACATAGAGTACCC 720  
 QY 242 AsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPhe 261  
 DB |||||  
 721 AACATGGAAGAGATGCGAGATCAGCTGAGTCCGCTACCTTGTCTCCAGAAAGAAATTT 780  
 QY 262 AspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLys 281  
 DB |||||  
 781 GATAGATATTTCTGGCTATTTGGTGTGTCCAAAAGCTTGAAACAACTCCCAAGTGTGTA 840  
 QY 282 IleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThr 301  
 DB |||||  
 841 ATTCTTAAGATTTCTATATGAGAAATATGATGATCTGAGGAAATTAATCATGTTTCA 900  
 QY 302 SerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAla 321  
 DB |||||  
 901 TCCCTATGTGGAAACAGAGAGGCGACATTCATTCCTTAACCAAGGTACAGCA 960  
 QY 322 AsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle 341  
 DB |||||  
 961 AATCCTTAAGTCACTTTAAGATGTCAAGAAATATGATGATGCTGAAGAGAAAGATCTCA 1020  
 QY 342 AspValIleAspLysGluLeuIleGlnProPheGlnIleLeuPheGlnGlyValGluTyr 361  
 DB |||||  
 1201 GATGTCATAGATTAAGAACTAATTCACCTTTTGAGATTCTAATTTGAAGAGGTTGAAT 1080  
 QY 362 IleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAspArg 381  
 DB |||||  
 1081 ATTCCCAAGAGCTGATGAGCTCCGAGGAGAAATATGCTGTGCTCATCTCTAGATGCG 1140  
 QY 382 SerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGluAsp 401  
 DB |||||  
 1141 TCCCAAGACTCGCCCTGCAAGATAGTGTATCTCACTGAATTAATTAATCCAGTGAAGAT 1200  
 QY 402 AspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeuIle 421  
 DB |||||  
 1201 GATGTTATGAAAGAGAGAGACTCATGAGTCAAGCTGATTCGTGAGCCCACTAATT 1260  
 QY 422 IleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhePro 441  
 DB |||||  
 1261 ATCTATGAAGAAACAGACATCTGATTAATATCATGACATCTTTCATGATTTTTCCTCC 1320

QY 442 GlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThrGlyPheArg 461  
 DB |||||  
 1321 CAAAGTCAAGAGAGAAATTAAGATTTATTTTGGCTCTGAATGCAAAACAGGTTTCCGT 1380  
 QY 462 HisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGlyGly 481  
 DB |||||  
 1381 CATTATATCAAAATTAATCATCTATTATTAAAGAAAGCAATATTAACGATCCAGTGTGG 1440  
 QY 482 LeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSerGly 501  
 DB |||||  
 1441 CTGCTGCTCCAGAGATTTCAAGTCTCTATCAAAAGGAGATAGCAATTAACCGTGGT 1500  
 QY 502 GluTrpGluValLeuGluArgHisGlySerAsnIleGlnValAspGluValArgArgLeu 521  
 DB |||||  
 1501 GAATGGAGATTTCTGGCCGCGCATGATATATATCAAGTTGATGAAGTCAAGAGGCTG 1560  
 QY 522 ValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSerTyr 541  
 DB |||||  
 1561 GTATATTTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGTAGTGTGAGTTAC 1620  
 QY 542 ValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysLysIle 561  
 DB |||||  
 1621 GTAAATCCTGGAGAGGTGACAAAGCTGACGACCGGTCTCACTCATCTTGTGCTGATC 1680  
 QY 562 SerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysVal 581  
 DB |||||  
 1681 AGTCAGCATGTGACTTTTATTAAGATGATAGTAAACCAAGAAATCCACACTGTGTG 1740  
 QY 582 SerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrp 601  
 DB |||||  
 1741 TCCCTTTAACAAGCTATCAAGTCTGAAGATGACCAACTTCCAAACAAAGAAATTTTGG 1800  
 QY 602 AlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSer 621  
 DB |||||  
 1801 GCCACCAATTTTGGATTGAGAGTCCAGAGTCCCTTCTGATCTACCTCCAGAAATTTTCTCT 1860  
 QY 622 PheGluSerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGln 641  
 DB |||||  
 1861 TTGAAGATCTACTGATTTTATCATATGATGAGATGCTTACCAAGCTCAAGATTTACAG 1920  
 QY 642 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGluLeuVal 661  
 DB |||||  
 1921 CCTGGAAGAAATATCTCACTGCTGTCTGTCATATATGTGTGCTCTCAGTGCAGTGTGG 1980  
 QY 662 AsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyr 681  
 DB |||||  
 1981 AATATCGGTTTAAAGAGATCAAGTATTTCCGCTTGAATACCTTCACTCTAGATTAT 2040  
 QY 682 ValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAla 701  
 DB |||||  
 2041 GTGGTTGTAGTATGACAAACAGGGGAGTCTGTCAACCGAGGGCTTAAATTTGAAGGCCGC 2100  
 QY 702 PheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeu 721  
 DB |||||  
 2101 TTTAAATTAATAAGGCTCAATGAAATGAAATGACATCAGGTGGAAGACTCCATATCTTA 2160  
 QY 722 AlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGly 741  
 DB |||||  
 2161 GCTTCTGATATGATTTTCATTTGATGATGATGCTGTGGGCAATCCAGCGCTGCTCATGGA 2220  
 QY 742 GlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAla 761  
 DB |||||  
 2221 GGATTAACCTCTCCGAGAGGCAATTATATCAGAGGTCAGATATCTTCAGGGTTGTATTGCT 2280  
 QY 762 GlyAlaProValThrLeuTrpIlePheTyrAspThrArgGlyTyrThrGluArgTyrMetGly 781  
 DB |||||  
 2281 GGGGCCCAAGTCACTCTGTGATCTTCTATGATACAGATACAGAAACGTTATATGGGT 2340  
 QY 782 HisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLys 801  
 DB |||||  
 2341 CACCTGACCAAGATGAAGAGGGCTATTAAGATCTGAGATCTGTGGCCATGCAAGCAAGAA 2400  
 QY 802 PheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHis 821



Db	2401	TTCCCTCTGAACCAATCGTTACTGCTCTTACATGTTTCCTGGAGAAATGTCAT	2460
Qy	822	PhealanhistrSerileuLeuSerPheleuValAlaGlaGlyLysProTyAspIleu	841
Db	2461	TTTGACATACAGATATATTACTGAGCTTTTATGAGAGCGCTGGAAAGCCATATGATTTA	2520
Qy	842	GlnIleTyProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyGlu	861
Db	2521	CAGATCTATCTCTCAGAGAGACACAGCATTAAGAGTTCTGTAATGGGAGAAATTATGAA	2580
Qy	862	LeuHisLeuLeuHisTyFleuGlnGluAsnLeuGlySerArgIleAlaLeuIleVal	881
Db	2581	CTGCATCTTTTGGACTACTCTTCAGAAACCTTGATACGATATTGCTGCTTAAAGTG	2640
Qy	882	Ile 882	
Db	2641	ATA 2643	
RESULT 8			
ABK83327	ID	ABK83327 standard; cDNA; 4829 BP.	
XX	AC	ABK83327;	
XX	DT	12-AUG-2002 (first entry)	
XX	DE	cDNA encoding human DPRP-1 splice variant #3.	
XX	XX	Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;	
XX	XX	DPP1V; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;	
XX	XX	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;	
XX	XX	heart failure; hypertension; urinary retention; osteoporosis; cancer;	
XX	XX	ulcer; allergy; cancer; psychotic disorder; neurological disorder;	
XX	XX	metabolic disorder; gene; ss.	
OS	XX	Homo sapiens.	
XX	PN	MO200231134-A2.	
XX	PD	18-APR-2002.	
XX	XX	12-OCT-2001; 2001WO-US31874.	
PF	XX	12-OCT-2000; 2000US-240117P.	
PR	XX	(FERR ) FERRING BV.	
XX	PA	Qi S, Akinsanya KO, Riviere PJ, Junien J;	
XX	PI	WPI; 2002-444178/47.	
XX	DR	P-PSDB; ABG61596.	
PT	XX	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding	
PT	XX	the proteins, useful for treating e.g. fungal, bacterial, protozoan and	
PT	XX	viral infections, cancers, allergies, neurological disorders, or pain	
XX	PS	Disclosure; Page 65-66; 11pp; English.	
XX	XX	The present invention relates to the isolation of novel human serine	
CC	XX	proteases referred to as dipeptidyl peptidase IV (DPP1V)-related	
CC	XX	proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)	
CC	XX	and nucleic acids encoding them are useful for treating infections	
CC	XX	such as fungal, bacterial, protozoan and viral infections, particularly	
CC	XX	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),	
CC	XX	pain, diabetes, precocious puberty, infertility, obesity, anorexia,	
CC	XX	bulimia, Parkinson's disease, acute heart failure, hypotension,	
CC	XX	hypertension, urinary retention, osteoporosis, angina pectoris,	
CC	XX	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,	
CC	XX	psychotic and neurological disorders (e.g. anxiety, dementia, or	
CC	XX	schizophrenia), and dyskinasias. These may also be used in discovering	

CC	therapeutic agents for the treatment of reproductive, inflammatory and
CC	metabolic disorders. ABK83322-ABK83343 encode human DPPF proteins.
XX	
XX	Sequence 4829 BP; 1466 A; 886 C; 1017 G; 1460 T; 0 other;
SQ	
Alignment Scores:	
Pred. No.:	0
Score:	4680.00
Percent Similarity:	99.77%
Best Local Similarity:	99.77%
Query Match:	99.57%
DB:	24
Gaps:	0
US-10-070-454-1 (1-882) x ABK83327 (1-8829)	
QY	1 MetAlaAlaIaMeGluThrGluGlnGluValGluIlePheGluThrAlaAspCys 20
Db	214 ATGCAGCGACGATGAAACAGAAACAGCGCGGTGTGAATATTTGAAATCGCGGACTGT 273
QY	21 GluGluuAnIleGluSerGlnAspArgProLysGluProPheTyrValGluATyr 40
Db	274 GAGGAGAAATATTGAATCACAGGATCGGCTAAATTGGACCTTTTATATGTGGAGCGGTAT 333
QY	41 SerTPSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db	334 TCCTGGAGTCACGCTTAAAAAGCTGCTGTCCGATACAGAAAATATCATGCTCATGATG 393
QY	61 AlAlvSAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db	394 GCTAAGGCGACCAATGATTTCAATGTTTGGAAGAGAGAAATGATCCAGATGACCTCATTTCA 453
QY	81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db	454 GACAGAAATCTATTACCTTGCCATGTCTGGTAGAAACAGAAATACTACTGTTTATCT 513
QY	101 GluIleProLysThrIleAsnArgAlaValIleuMetLeuSerTyrLysProLeuLeu 120
Db	514 GAAATTTCCAAAACATATCATATGAGACAGCATTTATATGCTCTTGGAAGCCTCTTTTG 573
QY	121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db	574 GATCTTTTTCAGGCAACACTGGACTATGGAATGATATCTCGAAGAAAGAACTATTAGA 633
QY	141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db	634 GAAAGAAAACGATTTGAACAGTCGGAATTTGCTTTCTTAGATATTATACCAAGGAAGTGA 693
QY	161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db	694 ACATTTCTGTTTCAAGCCCGTAGTGAAATTTATACGTAAGAAATGAGAGGCCACAAGGA 753
QY	181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db	754 TTTTAGCAACAACCTTTAAGGCCCAATCTAAGGGAACATGTTGTCCCAACATACGATG 813
QY	201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
Db	814 GATCCAAAATATTAGCCCTGCTGATCCAGACTGGATTTGCTTTATACATAGCAACGATATT 873
QY	221 TrpIleSerAsnIleValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
Db	874 TGGATATCTTAACATCGTAAACAGAGAAAGAGAGCTCACTTATGTGCAACATGAGCTA 933
QY	241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db	934 GCCAACATGGAAGAGATCCAGATCAGCTGGAGTGCCTACTTTGTTCTCCAAAGAAAGAA 993
QY	261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280
Db	994 TTTGATAGATATTCTGGCTATTGTGGTGTCCTCAAAAAGCTGAACAACATCCCAATGGTGTGT 1051
QY	281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300

Db 1054 AAAATCTTAGAATTCATATGAGAAATGATGATCTGAGTGGAAATTATTCATGTT 1113  
 Qy 301 ThrsrPmewtleuGluthrArxArxglAlaAspserPheArgTyrrProlyseThrGlyThr 320  
 Db 1114 ACATCCCTATGTTGGAAACAGAGGGAGATTCATTCGGTATCTTAAACAGGTACA 1173  
 Qy 321 AlaAspProLyValThrPheLyseSerGluIleWecIleAspAlaGluGlyArgIle 340  
 Db 1174 GCAAATCCTAAAGTCATTTTAAGATGTCAGAAATATGATTTGATGTCGAAAGGAGTCC 1233  
 Qy 341 IleAspValIleAspLyseGluLeuIleGlnProheGluIleLeuPheGluGlyValGlu 360  
 Db 1234 ATATAGTCATAGATTAAGAACTAATTCACCTTTTGATTCATTTTGAAAGGAGTTCAA 1293  
 Qy 361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLyseTyrrAlaTPSerIleLeuLeuAsp 380  
 Db 1294 TATATGCGACAGCTGGATGAGCTCTGAGGAAATATGCTTGCTGATCCATCTTAAT 1353  
 Qy 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
 Db 1354 CGCTCCGAGACTCGCTCAAGATAGTGTGATCTCAGCTGAATATTATCCAGTAGAA 1413  
 Qy 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420  
 Db 1414 GATGATGTTATGGAAGGACAGAGACTCATGAGTCAGGCTGATTTCTGACGCGACTA 1473  
 Qy 421 IleIleTyrrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440  
 Db 1474 ATTAATCTATGAGAAACACAGACATCTGATTAATATCCATGACATTTTCATGTTTTT 1533  
 Qy 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysAlaSerThrGlyPhe 460  
 Db 1534 CCCCAAGTCAAGAAAGAAATGAGTTATTTTGGCTGTAATGAAAACGGTTTC 1593  
 Qy 461 ArgHisLeuTyrrLyseIleThrSerIleLeuLyseGluSerLyseTyrrLyseArgSerSerGly 480  
 Db 1594 CGTCATTTATACAAATTTACATCTATTTTAAAGAAACAAATTAACGATCCACGAGGT 1653  
 Qy 481 GlyLeuProAlaProSerAspPheLyseAspProIleLyseGluGluIleAlaIleThrSer 500  
 Db 1654 GGGCTGCTGCTCCAGATGATTTCAAGTCTCTTCAAGAGAGATGCAATTAACCACT 1713  
 Qy 501 GlyLeuTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520  
 Db 1714 GGTGAATGGGAAGTCTTGGCGGCAATGATCATATACCAAGTTGATGAAGTCAGAAAG 1773  
 Qy 521 LeuValIleTyrrPheGluGlyThrLyseAspSerProLeuGluHisHisLeuTyrrValIleSer 540  
 Db 1774 CTGGTATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGTAGTACGTACGT 1833  
 Qy 541 TyrValAsnProGluValIleThrArgLeuThrAspArgGlyTyrrSerHisSerCysCys 560  
 Db 1834 TACGTAAATCTGGAGAGGTACCAAGGCTGACTACCGTGGCTACTCACTTTCTGCTGC 1893  
 Qy 561 IleSerGlnHisCysAspPhePheIleSerLyseTyrrSerAsnGlnLyseAsnProHisCys 580  
 Db 1894 ATCAGTCAGACCTGTGACTTTTATAGTAAGTATGTAACCGAAGAAATCCACACTGT 1953  
 Qy 581 ValSerLeuTyrrLyseLeuSerSerProGluAspAspProThrCysLyseThrLyseGluPhe 600  
 Db 1954 GTGTCCTTTACAGACTATCAAGCTCTGAGATGACCCCAACTTCCAAACCAAGAAATTTT 2013  
 Qy 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrrThrProProGluIlePhe 620  
 Db 2014 TGGGCGCACCTTTTGGATTCAGCGAGGCTCTTCTGCTATACTCTCCCAAGAAATTTTC 2073  
 Qy 621 SerPheGluSerThrThrGlyPheThrLeuTyrrGlyMetLeuTyrrLyseProHisAspLeu 640  
 Db 2074 TCTTTTGAAGTACTACTGATTTTACATTTGATGAGATGCTTCAAGGCTCAGATCTA 2133  
 Qy 641 GlnProGlyLyseLyseTyrrProThrValLeuPheIleTyrrGlyGly--ProGlnValGln 660  
 Db 2134 CAGCTGGAAAGAAATATCTACTGTCTTTCATATATGATGATGCTTCCAGGTCAGT 2193

Qy 660 euValAsnAsnArgPheLyseGlyValLyseTyrrPheArgLeuAsnThrLeuAlaSerLeuG 680  
 Db 2194 TGGTGAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTAGCTTCTAG 2253  
 Qy 680 LyTyrrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysePheGluG 700  
 Db 2254 GTTATGCTGTTGTAGTATACACAACAGGGATCTGTCAACGAGGCTTAAATTTGAAG 2313  
 Qy 700 LyAlaPheLyseTyrrLyseMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnT 720  
 Db 2314 GCGGCTTTAAATATAAATGGGTCAAAATAGAAATTTGACAGATCAGGTGAGAGACTCCAA 2373  
 Qy 720 TyrLeuAlaSerArgTyrrAspPheIleAspLeuAspArgValGlyIleHisGlyTPSerT 740  
 Db 2374 ATCTAGCTTCTCGATATATATTCATTTGACTTAGATCGTGGGCAATCCACGCTGGCTCT 2433  
 Qy 740 TyrrGlyGlyTyrrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI 760  
 Db 2434 ATGAGAGATACCTCTCCCTGATGGCATTTATGCAAGAGTCAAGATATCTTCAGGCTTGT 2493  
 Qy 760 leAlaGlyAlaProValThrLeuTrpIlePheTyrrAspThrGlyTyrrThrGluArgTyrrM 780  
 Db 2494 TTGCTGGGGCCCGAGTCACTCTGTGATCTTCTATGATACAGATACAGGAACGTTATA 2553  
 Qy 780 etGlyHisAspAspGlnAsnGluGlnGlyTyrrTyrrLeuGlySerValAlaMetGlnAlaG 800  
 Db 2554 TGGGTCACTCCGACACAGATACACAGGCTATTTACTTAAAGATCTGTGGCCATGCAAGCAG 2613  
 Qy 800 LyLysePheProSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnV 820  
 Db 2614 AAAAGTTCCTCTGAAACCAATGCTTACTGCTCTTACATGCTTTCTGATGAGAAATG 2673  
 Qy 820 aHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLyseProTyrrA 840  
 Db 2674 TCCATTTTGCACATACCAAGTATATTAATGATGATTTTATGAGGCTGAAAGCCATATG 2733  
 Qy 840 sPheGlnIleTyrrProGlnIleArgHisSerIleArgValProGluSerGlyGluHisST 860  
 Db 2734 ATTTACAGATATATCTCCAGAGAGACACACATTAAGAGTCTGAAATCGGAGAAACAT 2793  
 Qy 860 TyrrGluLeuHisLeuLeuHisTyrrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuL 880  
 Db 2794 ATGAACGCAATCTTTTGACACTTCTCAAGAAACCTTGGATCAGATTTGCTGCTTAA 2853  
 Qy 880 ysValIle 882  
 Db 2854 AAGTGATA 2861  
 Db  
 RESULT 9  
 ABK83332  
 ID ABK83332 standard; cDNA; 4685 BP.  
 XX  
 AC ABK83332;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DPP-1 splice variant #8.  
 XX  
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;  
 KW DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KW dykinegia; reproductive disorder; inflammatory disorder;  
 KW metabolic disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W0200231134-A2.  
 XX  
 PD 18-APR-2002.  
 XX

PF 12-OCT-2001; 2001WO-US31874.  
XX  
PR 12-OCT-2000; 2000US-240117P.  
XX  
PA (FERR) FERRING BV.  
XX  
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;  
XX  
XX MPI: 2002-444178/47.  
DR P-PSDB; AB661601.  
XX  
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
PT viral infections, cancers, allergies, neurological disorders, or pain  
PT  
PS Disclosure; Page 75-76; 113pp; English.  
XX  
XX The present invention relates to the isolation of novel human serine  
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
CC and nucleic acids encoding them are useful for treating infections  
CC such as fungal, bacterial, protozoan and viral infections, particularly  
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
CC biliary, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, osteoporosis, angina pectoris,  
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
CC schizophrenia), and dyskinesias. These may also be used in discovering  
CC therapeutic agents for the treatment of reproductive, inflammatory and  
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.  
XX  
SQ Sequence 4685 BP; 1430 A; 853 C; 991 G; 1411 T; 0 other;  
Alignment Scores:  
Pred. No.: 0 Length: 4685  
Score: 4385.50 Matches: 834  
Percent Similarity: 94.56% Conservative: 0  
Best Local Similarity: 94.56% Mismatches: 1  
Query Match: 93.31% Indels: 48  
DB: 24 Gaps: 1  
US-10-070-464-1 (1-882) x ABK83332 (1-4685)  
QY 1 Metalaalaametaugluhthrgluhneuglyvalgluilephnegluhthralaapcy 20  
DB 214 ATGCAGACGACGAAATGGAACAGACAGCTGGGTGATGATTTGAACTCGGACTGT 273  
QY 21 GluGluAenllEGuSerGluAapArgProlySleuGluProPhetYrValGluArgTyr 40  
DB 274 GAGGAGAAATATTGAATCACAGATCGGCTTAAATTGAGCTTTTATTTGAGCGGTAT 333  
QY 41 SERTPSEGLHLEUYSLEULEUAAASPThrArgLysTyrHisglYTYrMetMet 60  
DB 334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCAGATACAGAAATATCATGGCTCATGATG 393  
QY 61 AAlaYsAlaProHisaPheMetPheValYsArgAsnAspProAspGlyProHisSer 80  
DB 394 GCRAAGCAACACATGATTTGATTTGAGAAAGAAATGATCAATGAGTGAATCACTCA 453  
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAenThrLeuPheTyrSer 100  
DB 454 GACAGATCATTAATCCCTGCTGATGCTGGTGAAGAACAGAAATATCATCTGTTTATCT 513  
QY 101 GluIleProLysThrIleAsnArgAlaValLeuMetLeuSerTyrLysProLeuLeu 120  
DB 514 GAAATTCCTCAAAATCATCATTAAGACGAGCTTATGCTCTTGGAGGCTCTTTTG 573  
QY 121 AspLeuPheGlnIleThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuArg 140  
DB 574 GATCTTTTTCAGGCAACATGAGTGAATGATGATTTCTGAGAAAGAACTATTAGA 633

QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisglNglYSerGly 160  
DB 634 GAAAGAAAACGATGTGAACAGTCGAAATGCTTCTTACGATTAATACAGAGAGTGA 693  
QY 161 ThrPheLeuPheGlnIleArgLysSerGlyIleTyrHisValYsAspGlyGlyProGlnGly 180  
DB 694 ACATTTCTGTTTCMAACCGGTAGTGAATTTATCATCTAAATAATGAGGCGCAAGAGA 753  
QY 181 PheThrGlnIleProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
DB 754 TTTACGCAACACCTTTAAGGCCCATCTAGTGAAACTAGTTGCTCCACATACGATG 813  
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220  
DB 814 GATCCAAATTAATGACCTGCTGATTCAGACTGATGCTTTTATACATAGCAACGATATT 873  
QY 221 TrrPLeSerEnllEValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240  
DB 874 TGGATATCTAACATCGTAACCAAGAAAGAAAGAGACTCATTAATGACCAATAGCTGA 933  
QY 241 AlaAsnMetGluGluAspAlaArgSerArgIleValAlaThrPheValLeuGlnGlu 260  
DB 934 GCCAACTGGAGAAAGATGCGAATCAGCTGAGAGTCCCTACCTTGTCTCCAAAGAA 993  
QY 261 PheAspArgTyrSerGlyTyrTrrPTrCysProLysAlaGluThrThrProSerGlyGly 280  
DB 994 TTTGATGATATTCCTGGCTATTGGTGTGTCCAAAAGCTGAACAACTCCAGTGGTGCT 1053  
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisVal 300  
DB 1054 AAAATTTCTTAATTAATCTTAATGAAGAAATGATGATTTGAGGAGAAATATTATCATGTT 1113  
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
DB 1114 ACATCCCTTAATGTTGAAGAAACAGAGGCGAGATTCATCCGTTATCTTAAACAGATGCA 1173  
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340  
DB 1174 GCAAACTCTAAAGTCACTTTTAAGATGTCAGAAATTAAGATTTGAGTGAAGAAAGATC 1233  
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGluValGlu 360  
DB 1234 ATGATGTCATGATGATTAAGAACTAATTTACCTTTTGAGATTCTAATTTGAAGAGATTGA 1293  
QY 361 TyrIleAlaArgAlaGlyTrrThrProGluGlyLysTyrAlaTrrPTrIleLeuLeuAsp 380  
DB 1294 TATATTGCCAGAGCTGGAATGAGCTCTGAGGAAATATGCTTGATCTTACTAGAT 1353  
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
DB 1354 CCTCCCAAGACTCGCTACAGATGATGATCTCACTGAATTAATTTATCCCACTAGAA 1413  
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420  
DB 1414 GATGATGTTATGAAAGGACAGACATCATTTGATGATGATGATCTGATCTGAGCGCACA 1473  
QY 421 IleIleTyrGluGluThrThrAspIleTrrPLeAsnIleHisAspIlePheHisValPhe 440  
DB 1474 ATTATTAAGAAAGAAACACAGACATCTGATTAATATCAAGACATCTTCAATGTTTTT 1533  
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysTyrThrGlyPhe 460  
DB 1534 CCCCAAGTCAAGAAAGAAATGATTTATTTTGTCTGAAATGCAAAACAGGTTTC 1593  
QY 461 ArgHisLeuTyrTyrIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480  
DB 1594 CGTCATTTATACAAATTTACATCTATTTTAAAGAAAGCAAAATTAACATCAAGTGT 1653  
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500  
DB 1654 GGGCTGCTGCTCCAAAGTGAATTTCAAGTGCCTATCAAGAGAGATAGCAATTAACAGT 1713  
QY 501 GlyGluTrrPgluValLeuGlyArgHisglYSerEnllEglnValAspGluValArgArg 520

```

Db 1714 GGTGAATGGAGAGTCTTGCGCGCATGATCTAATATCCAAAGTTGAGAAAGTGAAGG 1773
Qy 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLysLeuTyrValLysSer 540
Db 1774 CTGGTATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGATGAGTCACTGAGTCACT 1833
Qy 541 TyrValAspProGlyValValThrArgLeuThrAspArgLysTyrSerHisSerCysGly 560
Db 1834 TACGTAATCTCGAGAGGTGACAAAGGCTGACTGACCTGGCTACTCACTTCTTGCTGC 1893
Qy 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db 1894 ATCAGTCAGCACTGATCTTTATTAAGTAAGTATTAATACCAAGAAATCCACACTGT 1953
Qy 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db 1954 GTGCCCTTTTCAAGCTATCAAGTCTGAAAGTGAAGCCCAACTGGCAAAAACAAAGAAATTT 2013
Qy 601 ThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe 620
Db 2014 TGGGCACACATTTGGATTCACT----- 2036
Qy 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db 2036 ----- 2036
Qy 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db 2037 -----CCTCAGGTGCACTTG 2051
Qy 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuLysSerLeuGly 680
Db 2052 GTGAATATCGGTTTAAAGAGTCAAGTATTTCCCTGAAATACCTTGAAGCTCTCTAGGT 2111
Qy 681 TyrValValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGly 700
Db 2112 TATGTGGTTGATGATGATGACACAGGGAGTCTGTCAAGGGCTTAAATTTGAAGGC 2171
Qy 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db 2172 GCCCTTAAATATTAATGCGTCAATAGAAATGACATCGATCGAGGAGGAGCATCCAAATAT 2231
Qy 721 LeuLysSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
Db 2232 CTACCTCTCGATATGATTTCACTTGAATGCTGAGGAGATCCACAGGCTGGGCTCAT 2291
Qy 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db 2292 GGAAGATACCTCTCCCTGATGGCATTTATGCAAGAGTCAGATATCTTCAAGGCTTCTATT 2351
Qy 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db 2352 GCTGGGGGCCAGACACTCTGTGATCTTATGATACAGATACACGGAACGTTATATG 2411
Qy 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db 2412 GGTACACCTGACCGAATGAACAGGGCTATTACTTAAGATCTGGGCCATCAAGCAAA 2471
Qy 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db 2472 AAGTTCCCTCTGAACCAATCGTTTACTGCTTTACATGTTTCTCGATGAGAAATGTC 2531
Qy 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLysProTyrAsp 840
Db 2532 CATTTTCACATACACAGTATATTACTGAGTTTATGAGAGGCTGGAAACCATATGAT 2591
Qy 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db 2592 TTACAGATCTATCTCTCAGAGAGACACAGCATTAAGAGTTCTGAAATCGGGAACATTA 2651
Qy 861 GluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys 880

```

```

Db 2652 GAAGTCATCTTTTGACATACCTTCAAGAAAACCTTGATACAGTATGCTGCTAATA 2711
Qy 881 ValIle 882
Db 2712 GTGATA 2717
RESULT 10
ID ABK83331
ABK83331 standard; cDNA; 4676 BP.
XX
AC ABK83331;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPP-1 splice variant #7.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyslexia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO20021134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX WPI; 2002-444178/47.
XX DR P-PSDB; ABG61600.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
PS Disclosure; Page 72-73; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPP proteins.
XX
SQ Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0 Length: 4676
Score: 4385.00 Matches: 831
Percent Similarity: 94.22% Conservative: 0
Best Local Similarity: 94.22% Mismatches: 1
Query Match: 93.30% Indels: 51
DB: Gaps: 1

```

US-10-070-464-1 (1-882) x ABK83331 (1-4676)

QY 1 MetAlaAlaMetGluThrGluGluLeuGlyValGluIlePheGluThrAlaAspCys 20  
DB 214 ATGGAGCGCAGCATGGAAAACAGAACGCTGGGTGGAGATATTGAACTCGGACTCT 273  
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40  
DB 274 GAGGAGAAATTGAAATCACAGATCGGCTTAAATTGGAGCCTTTTATGTGAGCGGTAT 333  
QY 41 SerTyrSerGluLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60  
DB 334 TCTCGAGTCAGCTTAAAGAGCTGCTTCCGATACCAAGAAATATCATGGCTACATGATG 393  
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80  
DB 394 GCTAAGGACCAACATGATTCATGTGTTGAAAGAGAAATGATCCAGATGAGCCTCATTTA 453  
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyLysAsnArgGluAsnThrLeuPheTyrSer 100  
DB 454 GACAGATCTATTACCTTCCATGTGTGTGAGAACAGAAATATCATGCTTTTATCTT 513  
QY 101 GluIleProLysThrIleAsnArgAlaValLeuMetLeuSerTyrLysProLeuLeu 120  
DB 514 GAAATTTCCAAACATCATCATGAGCAGCAGCTTATGCTCTCTTGGAAAGCCTCTTTTG 573  
QY 121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140  
DB 574 GATCTTTTTCAGGCAACCTGAGCATATGGAATGTATCTCGAAGAAAGAACCTATTAA 633  
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnLysSerGly 160  
DB 634 GAAAGAAACGATTTGAGACAGTCGGAATTTGCTTTCATGATTTATCCAAAGAACTGA 693  
QY 161 ThrPheLeuPheGluAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180  
DB 694 ACATTTCTGTTCAGACCCGAGTGTGAATTTATCATGTAAGATGAGAGGCGACAAAGA 753  
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
DB 754 TTTAGCCAAACACCTTTAAAGCCCAATCTAGTGGAACTAGTGTCTCCAACTACCGAATG 813  
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220  
DB 814 GATCCAAATTTATGCTGCTGCTGATCCAGCTGGAATTTGCTTTATCATGCAACGATAT 873  
QY 221 TrrPileSerAsnIleValThrArgGluGluArgLeuThrTyrValHisArgGluLeu 240  
DB 874 TGGATATCTAAACATCGTAACAGAGAAAGAGAGACTCATGATGTGCACATGAGCTA 933  
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260  
DB 934 GCCAACATGGAAAGAAAGATCCAGATCCAGTGGAGTGCATCTTGTCTCCAAAGAA 993  
QY 261 PheAspArgTyrSerGlyTyrTrrPrrCysProLysAlaGluThrThrProSerGlyGly 280  
DB 994 TTTGATAGATATTCTGGCTATTGGTGTGTCCAAAGAGCTGAAACAACTCCAGTGTGT 1053  
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300  
DB 1054 AAAATTTCTTAAGATCTATATGAAGAAATGATGATCTGAGGTGGAATTTATTCATGTT 1113  
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
DB 1114 ACATCCCTCATTTTGGAAACAAAGAGGCGAGATTCATCTCGTTATCTTAAACAGGTACA 1173  
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340  
DB 1174 GCAAAATCCTAAGTCACTTTTAAGATGTGAGAAATATGATGATGCTGAAGAGAGATC 1233  
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360  
DB 1234 ATAGATGTATGATTAAGAACTAATTCACCTTTTGAGATTTCTATTGAAAGAGATTGA 1293

QY 361 TyrIleAlaArgAlaGlyTrrPrrProGluGluLysTyrValAlaTrpSerIleLeuLeuAsp 380  
DB 1294 TATATTTCCAGAGCTGAGTGAAGCTCTGAGGAGAAATATGCTTGGTCCATCTCAGAT 1353  
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
DB 1354 CGCTCCCAAGCTCGCTACAGATAGTGTGATCTCACCCTGAATTTATTTATCCAGTAGA 1413  
QY 401 AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu 420  
DB 1414 GATGATGTATGGAAGGAGAGACTCATTTGAGTCACTCGATCTCTGTGAGCCACATA 1473  
QY 421 IleIleTyrGluGluThrThrAspIleTrrPrrIleAsnIleHisAspIlePheHisValPhe 440  
DB 1474 ATTATCTATGAAGAAACAAACAGACATCTGGAATTAATTCATGACATCTTTCAATGTTT 1533  
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGlyCysLysThrGlyPhe 460  
DB 1534 CCCCAAGTCAAGAAAGAGAAATTTGATTTATTTTGGCTCGAATGCAGAAACAGTTTC 1593  
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480  
DB 1594 CGTCATTTATACAAATTTACATCTATTTTAAAGAAAGCAAAATATTAACGATCCAGTGT 1653  
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500  
DB 1654 GGGCTGCTGCTCCAAAGATTTCAAGTGTCTTATCAAAAGAGAGATACCAATTTACAGT 1713  
QY 501 GlyIuTrrPrrGluValLeuGlyYArgHisGlySerAsnIleGlnValAspGluValArgArg 520  
DB 1714 GGTGAATGGGAAGTTCTTGGCCGCGATGATCTTAATATCCAGTTGATGAATCGAAG 1773  
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSer 540  
DB 1774 CTGGTATTTTGAAGGACCAAGACCTCCCTTTAGAGCATCACCTGTACATGACT 1833  
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560  
DB 1834 TACGTAATCTCGAGAGGTGACAGGCTGAGTCCGCTGACTACATCTTGTGCTGC 1893  
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580  
DB 1894 ATCAGTACAGCATGTGAGCTCTTTATAGTAAGTATGTAACAGAGATCCACACTGT 1953  
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600  
DB 1954 GTGTCCCTTTACAAAGCTATCAAGTCTGAGAGATGACCCCAACTTGCAGAAAGCAATTT 2013  
QY 601 TrrPrrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620  
DB 2014 TGGGCCACCATTTTGGATTCAGAGGCTCTCTCTGACTATATCTCTCAGAAATTTTC 2073  
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640  
DB 2074 TCTTTTGAAGTACACTGAGATTTTACATTTGATGAGATGCTCTACAGCTCATATGTA 2133  
QY 641 GlnProGlyLysGlyTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660  
DB 2134 CAGCTCGAAGAAATATCTCATCTGTGCTTCATATATGTGGTGTGCG----- 2180  
QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680  
DB 2180 ----- 2180  
QY 681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700  
DB 2180 ----- 2180  
QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720  
DB 2181 -----GGTCAATATGAATAATTGACATCGATCGGTGAAGGATCCCAATAT 2222

QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740  
DB 2223 CTAGCTTCTCATATGATGATTTTATGACTTACATGCTGGGACATCCAGCGCTGGCTCAT 2282  
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760  
DB 2283 GGAGGATACCTCTCCCTGATGCGATTATGACAGAGGTGAGATATCTTCAAGGTGCTATT 2342  
QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780  
DB 2343 GCTGGGGCCCAAGTCACTCTGTGATCTTATGATACAGATACACGGAACGTTATATG 2402  
QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800  
DB 2403 GGTACCTGACCAAGATGAACAGGGCTATTACTTAGATCTGGCCATGCAAGCAGAA 2462  
QY 801 LysBheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820  
DB 2463 AAGTTCCCTCTGAACCAATCGTTTACTGCTCTTACATGGTTCTCGATGAGATGTC 2522  
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840  
DB 2523 CATTTTGACATACAGATATATTACTGAGTTTGTAGAGGGCTGGAAAGCCATATGAT 2582  
QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlnSerGlyGluHisTyr 860  
DB 2583 TTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATTCGGAATCGGAGAACATTAT 2642  
QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys 880  
DB 2643 GAATGCATCTTTTGCATCACTTCAAGAAACCTTGATCAAGATTCGCTCTAAA 2702  
QY 881 ValIle 882  
DB 2703 GTGATA 2708  
RESULT 11  
ID ABN59774 standard; cDNA; 2842 BP.  
AC ABN59774;  
DT 28-JUN-2002 (first entry)  
XX Novel human coding sequence SEQ ID NO: 185.  
DE Human; anti-nausea; vulnerability; anti-inflammatory; immunomodulator;  
KW anti-infectivity; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KM neuroprotective; anti-parkinsonian; protein therapy; EST;  
XX expressed sequence tag; gene; ss.  
OS Homo sapiens.  
XX MO200222660-A2.  
XX 21-MAR-2002.  
XX 10-SEP-2001; 2001MO-US26015.  
XX 11-SEP-2000; 2000US-0659671.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Zhou P, Aundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX WPI; 2002-292408/33.  
XX P-PSDB; ABB97361.  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -  
XX Claim 1; SEQ ID NO 185; 509bp; English.

XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haemacopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a coding sequence of the  
CC invention.

SO Sequence 2842 BP; 857 A; 592 C; 635 G; 758 T; 0 other;

#### Alignment Scores:

Pred. No.:	Length:	Matches:	Mismatches:
Score:	0	2842	782
Percent Similarity:	418.00		
Best Local Similarity:	88.66%		
Query Match:	87.62%		
DB:	24	Gaps:	1

US-10-070-464-1 (1-882) x ABN59774 (1-2842)

QY 1 MetaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20  
DB 234 ATGCACACAGCAATGAAAGCAAGACAGCTGGGTGTGAGATATTGAACTGCGGACTGT 293  
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40  
DB 294 GAGGAGATATGGAATACAGAGATCGGCTTAATTTGAGCCCTTTTATGTTGACCGGTAT 353  
QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60  
DB 354 TCCTGAGTCAAGCTTAATAAAGCTGCTTCCGATACCAAGAAATATCATGCTATCATATG 413  
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80  
DB 414 GCTAAGCACACACATGATTTTATGTTGTAAGAGAAATGATCCAGATGACCTCATTTCA 473  
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100  
DB 474 GACGAATCTATTACTTGCATGCTGTGTAGAACAGAGAAATTAACATGTTTATTTCT 533  
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120  
DB 534 GAATTCACAAACTATCAATAGACAGCTTATGCTCTTGGAACTCTTTTG 593  
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140  
DB 594 GATCTTTTTCAGGCAACACTGACTGATGATGATTTCTCGAGAAAGAACTATTAGA 653  
QY 141 GluArgLysArgGlyIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160  
DB 654 GAAAGAAAGACATGGAACAGTCGGAATTTGCTTTACATTAATACCAAGAGAGTGA 713  
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180  
DB 714 ACAATTTCTGTTTCAAGCCGATGATGAAATTTATCACTAAAGATGAGGACCAAGGA 773  
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
DB 774 TTACGCAACACCTTTAAGGCCCAATCTAGTGAACCTAGTTGCCAATATCGGATG 833  
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220  
DB 834 GATCCAAATTAATGACCTGCTGATCCAGCTGAGATTTGCTTTATACATACCAACGATATT 893  
QY 221 TyrIleSerAsnIleValThrArgGluGluArgLysLeuThrTyrValHisAsnGluLeu 240  
DB 894 TGGATATCTTAACATCGTAACCGAAGAAAGAGACTCATTAATGCAATGAGCTA 953

QY 241 AlaasmeGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260  
 DB 954 GCCACATGGAAAGAGATGCCAGATCACTGAGTCCGACTTGTCTTCAGAAAGAA 1013  
 QY 261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGly 280  
 DB 1014 TTGGATAGATATCTGGCTATGGTGGTGCMAAGCTGAAACAACCTCCAGAGGTGGT 1073  
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300  
 DB 1074 AAAATTCTTGAATTCATATAGAAAGAAATGATCAATCTGAGGTGAATTAATTCATGTT 1133  
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
 DB 1134 ACATCCCTATGTTGAAACAGAGAGGCGAGATTCATCCGTTATCTTAACACGGATCA 1193  
 QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340  
 DB 1194 GCAATCTCTAAGTCACTTTTAAAGATGTCAGAAATTAATGATGATGCTGAAGGAAGATC 1253  
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360  
 DB 1254 ATGAGTGTCAATGATTAAGAACTAATTCACCTTTGAGATTCATTTGAAGAGATTGAA 1313  
 QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaTyrPserIleLeuLeuAsp 380  
 DB 1314 TATATTCACAGAGCTGATGAGTCTCGAGGGAATAATGCTTGCTTCATCTCTATGAT 1373  
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
 DB 1374 CGCTCCAGACTCCGCTCAAGATAGTGTGATCTCACCTGAATTAATTAATCCAGTAGAA 1433  
 QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420  
 DB 1434 GATGATGTATGGAAGGACAGAGACTATGATGACGCTGATTCGTGACCCACTA 1493  
 QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440  
 DB 1494 ATATATCATGAAGAAACAAGACATCGATTAATTCATGACATCTTTCATGTTTT 1553  
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460  
 DB 1554 CCCCAAGTCAAGAAAGAAATGAGTTATTTTGGCTCTGAAATCAAAACGGTTTC 1613  
 QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480  
 DB 1614 CGTCAATTAACAATAATTAATCAATTTTAAAGAAAGCAAAATTAACGATCCAGTGGT 1673  
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500  
 DB 1674 GGGCTGCTGCTCCAGAGATTTCAAGTGTCTTCAAAAGAGAGATAGCAATTAACAGT 1733  
 QY 501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520  
 DB 1734 GGTGAATGGGAAGTCTTGGCGGCGCATGATCTAATATCCAAAGTGTGTAAGTCAGAGAG 1793  
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSer 540  
 DB 1794 CTGGTATATTTGAAGGACCAAGACCTCCCTTTAAGCATCACCTGTAGTGATGACGT 1853  
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560  
 DB 1854 TACGTAAATCCTGAGAGGTGACAAAGCTGACTGACCGTGGCTACTCACAATTTTGGCTGC 1913  
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerArgGlnLysAsnProHisCys 580  
 DB 1914 ATCAGTCAGCACTGTGACTTCTTATAGTAAGTAAGTAACCGAAGAAATCCACACTGT 1973  
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600  
 DB 1974 GTGGCCCTTTAACAAGCTATCAAGTCCGTGAAGATGACCAACTTTCAAAACAAAGAAATTT 2033  
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620

DB 2034 TGGGCCACCATTTTGGATTTCAGAGGTCCTCTTCTCATACTATCTCCAGAAATTTTC 2093  
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640  
 DB 2094 TCTTTGAAGTCACTACTGATTTTACATGTATGAGGATGCTCTCAAGCCATGATCTA 2153  
 QY 646 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660  
 DB 2154 CAGCTGAAAGAAATATCTCTACTGCTGTCTCATATATGATGATGCTCTCAG----- 2204  
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680  
 DB 2204 ----- 2204  
 QY 681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700  
 DB 2204 ----- 2204  
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720  
 DB 2204 ----- 2204  
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740  
 DB 2204 ----- 2204  
 QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760  
 DB 2205 -----GTTGCTAATT 2213  
 QY 761 AlaGlyAlaProValThrLeuTyrPheThrAspThrGlyTyrThrGluArgTyrMet 780  
 DB 2214 GCTGGGGCCCAAGTCACTGTGGATCTTATGATACAGAGATACAGAGAACTTATATG 2273  
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800  
 DB 2274 GGTCACTGACCAAGATGAACAGGGCTATTACTTGAATCTGTGGCCATGCAAGCAAGAA 2333  
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820  
 DB 2334 AAGTTCCTCTGAAACCAAACTGTTACTCTCTTACATGATGTTCCGTGATGAAAGTGC 2393  
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840  
 DB 2394 CATTTGCACATACCAATATATTAATGAGTTTTTATGAGGCTGGAAGCCATATGAT 2453  
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyr 860  
 DB 2454 TTACAGATCTATCTCTGAGAGACACAGCATTAAGATTTCTGAATGGAGAAATTTAT 2513  
 QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880  
 DB 2514 GAACGTCATCTTTTGCATCACTTCAAGAAAACTTGATACAGATATGCTGCTAAAA 2573  
 QY 881 ValIle 882  
 DB 2574 GTGATA 2579  
 DB  
 RESULT 12  
 ID ABK63325  
 XX ABK63325 standard; cDNA, 4523 BP.  
 AC ABK63325;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DPP-1 splice variant #1.  
 XX  
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;  
 KW DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;



KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM	dyskinesia; reproductive disorder; inflammatory disorder;
KM	metabolic disorder; gene; ss.
XX	
OS	Homo sapiens.
PN	WO200231134-A2.
XX	
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US31874.
XX	
PR	12-OCT-2000; 2000US-240117P.
XX	
PA	(FERR ) FERRING BV.
XX	
PI	Qi S, Akinsanya KO, Riviere PJ, Juntien J;
XX	
DR	WPI: 2002-444178/47.
XX	
DR	P-PDB; ABG61594.
XX	
PT	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT	the proteins useful for treating e.g. fungal, bacterial, protozoan and
PT	vital infections, cancers, allergies, neurological disorders, or pain
PT	-
XX	
PS	Disclosure; Page 61-62; 113pp; English.
XX	
CC	The present invention relates to the isolation of novel human serine
CC	proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC	proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC	and nucleic acids encoding them are useful for treating infections
CC	such as fungal, bacterial, protozoan and viral infections, particularly
CC	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC	pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC	bulimia, Parkinson's disease, acute heart failure, hypotension,
CC	hypertension, urinary retention, osteoporosis, angina pectoris,
CC	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC	psychotic and neurological disorders (e.g., anxiety, dementia, or
CC	schizophrenia), and dyskinesias. These may also be used in discovering
CC	therapeutic agents for the treatment of reproductive, inflammatory and
CC	metabolic disorders. ABRK3322-ABRK3343 encode human DPRP proteins.
XX	
SQ	Sequence 4523 BP; 1384 A; 828 C; 940 G; 1371 T; 0 other;
Alignment Scores:	
Pred. No.:	0 Length: 4523
Score:	4092.50 Matches: 780
Percent Similarity:	88.44% Conservative: 0
Best Local Similarity:	88.44% Mismatches: 1
Query Match:	87.07% Indels: 102
DB:	24 Gaps: 1
US-10-070-464-1 (1-882) x ABRK3325 (1-4523)	
OY	1 MetAlAAlAAmEgJutHrGlugInleuNgIyValGIuIllePhEGlUThrAlAsPCys 20
DG	214 ATGGCAGCGACATGTGAACAAGAACAGTGGTGTTGAGATTATTGAAACTGCCGAGCTGT 273
OY	21 GlUGluAsnIleGlnSerGItnSPArPrOlySlenuGluPProPHetYyValGIUAITYr 40
Db	274 GAGAGAGAATTTGAAATCACAGATCGGCCTTAATTGGAGCCCTTTTATGTGGACGGTAT 333
OY	41 SetTrpSerGItnleuYslYslSeuleuAlaEPTrArGLyTYRHAGLYTYrMetMet 60
Db	334 TCCTGGAGTCAGCTTTAAAAAGCTGCTGCCGATACCAGAAAATATCATGCTCAATATG 399
OY	61 AlaLysAlaProHisAPheMetPheValIySArGaSnAPProAspGIYPProHisSer 80
Db	394 GCTAAAGCACCATGATTTTCATGTTTGGAAGAGAAATGATCCAGATGAGACCTCATTC 453
OY	81 AspArgIIleTyrrTYrLeuAlaMetSerGIyGiuaAnaGgluaErThrleuPhetyrSer 1000

Db	454	GACAGAAATTATTACCTTCCATGTCGTGTGAGAACAGAAATAATCACTGTTTATTCT	513
Qy	101	GIuIlleProIySThrIleaSnArGAlaIaValleuMeIleuSerTrpIySProIleu	120
Db	514	GAAATTTCCAAAACATATCATATGACAGACAGCTTAATGCTCTTTGGAAAGCTCTTTTG	573
Qy	121	AspIeuPheGlnAlaThrIleuAspTrgIyMeTrYSerArgIuGIuIleuLeuArg	140
Db	574	GATCTTTTTCAGGCAACAACCTGACTATGGAATGTATTCGAGAAAGAAACCTATTAAAG	633
Qy	141	GIuArgIyVAArgIleGIyThrAlaGIyIleAlaSerTrYAspTrhIseGIuIySergIy	160
Db	634	GAAAGAAAAACGATTTGGAACACATCGAAATTCCTTTACGATATTATCCACAGCAAGCTGA	693
Qy	161	ThrPheIeuPheGlnAlaGIySergIyIleTrhIseValIyAspGIyGIyProGIuGIy	180
Db	694	ACATTTCTGTTTCAAGCCCGTAGTGTGAATTTATCACTGTAAGATGAGAGGCCACAAAGA	753
Qy	181	PheThrgIuGlnProIeuArgProAsnIeuValGIuThrSerCyAsProAsnIleArgMe	200
Db	754	TTTACGCAACAACCTTTAAAGCCCATCTAGTGAACACTAGTTGTCCTCCAAACATACGATG	813
Qy	201	AspPheIySleuCyAsProAlaAspProAspTrpIleAlaPheIleHISerIeAspIle	220
Db	814	GATCCAAAATTAATGCGCTGCTGATCCAGCTGGATTCCTTTATACATAGCAACGATTAT	873
Qy	221	TrpIleSerAsnIleValThrArgGIuIuArgArgIeuThrTyValHISanGIuIeu	240
Db	874	TGGATATCTTAACATCTGTAACAGAGAAAGAAAGAACTCATTAATGTGCACATAGCTTA	933
Qy	241	AlaAsnMeGIuGIuAspAlaArgSerAlaGIyValAlaThrPheValIeuGIuGIu	260
Db	934	GCCAACTGCAAGAAAGATGCCAGATCAAGTGGAGTCGCTACTTTGTTCTCCAAAGAGA	993
Qy	261	PheAspAlaGTrYSergIyTrpTrpIyCysProIyValaGIuThrThrProserGIyGIy	280
Db	994	TTGTATAAATATTCGGCTATTTGGTGGTGTCCAAAAGCTGAAACAACCTCCAGTGGTGGT	1053
Qy	281	LysIleIeuArgIleIeuTrGIuGIuAsnAspGIuSergIuValGIuIleIleHISVal	300
Db	1054	AAAAATCTTAGAATTCATATGAGAAATAATGATGATCTAGAGTGAATAATTATTCATGTT	1113
Qy	301	ThrSerProMeIleuGIuThrArgArgAlaAspSerPheArgTrYProIySThrGIyThr	320
Db	1114	ACATCCCTTAATGTTGGAACAAGAGGCGCAAGTTCAATCCGTATCCATAAACAGAGTACA	1173
Qy	321	AlaAsnProIySValThrPheIyMeSergIuIleMeIleAspAlaGIuIyArgIle	340
Db	1174	GCAATATCTTAAGTCACCTTTTAAGATGTGCAGAAATAAAGATTGAATGCTGAAAGAGATC	1233
Qy	341	IleAspValIleAspIySleuIleuIleGlnProPheGIuIleuPheGIuGIyValaGIu	360
Db	1234	ATAGATGTCATAGATTAAGAACTAATTCCTTTTGAGATTCATATTGAAGGAAGTGTGA	1293
Qy	361	TyrIleAlaArgAlaGIyTrpThrProGIuGIyIySValaIyAspSerIleIeuIeuAsp	380
Db	1294	TATATTGCGAAGCTGATGATGATCTCCAGGGAATAATGCTGTGCATCTCACTTAAGT	1353
Qy	381	ArgSergIuThrArgIeuGlnIleValIeuIleSerProGIuIeuPheIleProValaGIu	400
Db	1354	CGCTCCAGACCTCCCTTACAGATAGTGTGTGATTCACCTGAATTAATTATCCAGTAGAA	1413
Qy	401	AspAspValMeGIuArgGIuArgIeuIleGIuIeuIleProAspSerValThrProIeu	420
Db	1414	GATATGTTATGAGAAAGCAGACAGCTAATGATCAGAGCCTGATTCGTATACGCCACTA	1473
Qy	421	IleIleTrGIuGIuThrThrAspIleTrpIleAsnIleHISAspIlePheHISValaPhe	440
Db	1474	ATTATCTATGAGAAACAACAGACATCTGATTAATATCCATGACATCTTTCAAGTTTTT	1533
Qy	441	ProGlnSerHISGIuGIuGIuIleGIuPheIlePheAlaSerGIuCySValSThrGIyPhe	460
Db	1534	CCCCAAAGTCACGAAGAGAAATGAATTAATTTTGGCTCTGATATGCAAAACAGGTTTC	1593

QY 461 ArgHisLeuTyrValIleThrSerIleLeuLysGluSerLysTyrTyrValArgSerSerCly 480  
 Db 1594 CGTATTATATACAAATTACATCTATTTTAAAGAAACCAATTAACGATCCAGTGT 1653  
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500  
 Db 1654 GGGCTGCTGCTCCAGGATTTTCAGAGTCTCTATCAAGAGAGATGACATTAACCGT 1713  
 QY 501 GlyLeuTyrGluValLeuGlyArgHisGlySerAsnIleGluValAspGluValArgArg 520  
 Db 1714 GGTGAATGGGAAGTCTTGCGCGGATGATCTAATATCCAAAGTTAGTAAGTCCAGAGG 1773  
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSer 540  
 Db 1774 CTGGTATATTTTGAAGGCACAAAGACTCCCTTTAGAGACATCACTGTACGTACGTACGT 1833  
 QY 541 TyrValAsnProGlyGluValIleThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560  
 Db 1834 TACGTAAATCTCGAGAGGTGACAAAGCTGACTGACCTGGCTACTCATCTTGTCTGC 1893  
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580  
 Db 1894 ATCAGTCAACACTGCTGACTCTCTTATAGTAAAGTATAGTAACAGAAATCCACACTGT 1953  
 QY 581 ValSerLeuTyrLysLysTyrProThrSerProGluAspAspProThrCysLysThrLysGluPhe 600  
 Db 1954 GTGGCTCTTTACAAAGCTATCAAGTCTTAAGATGACCCAACTGGAAAAACAAGAAATTT 2013  
 QY 601 TyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe 620  
 Db 2014 TGGGCGACCATTTTGATTCAGCAGGTCTCTCTCTGACTATACCTCCCGAAATTTTC 2073  
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640  
 Db 2074 TCTTTTGAAGTACTACTGAGTTTACATTTGATGAGATGCTCTCAACACCTCATGATCTA 2133  
 QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660  
 Db 2134 CAGCTGGAAAGAAATATCTCTACTGCTGCTCTCATATATGGTGGT----- 2178  
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680  
 Db 2178 ----- 2178  
 QY 681 TyrValValValIleAspAsnArgLysCysHisArgGlyLeuLysPheGluGly 700  
 Db 2178 ----- 2178  
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720  
 Db 2178 ----- 2178  
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740  
 Db 2178 ----- 2178  
 QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760  
 Db 2179 -----CGGT--GCTATT 2189  
 QY 761 AlaGlyAlaProValIleThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780  
 Db 2190 GCTGGGCGCCACAGTCACTCTGTGATCTTCTATGATACAGATACAGAAACGTTATATG 2249  
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800  
 Db 2250 GGTCACCTTGACCAAGATGAACAGGGCTATTAAGATCGAGCATCAAGACAGAA 2309  
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820  
 Db 2310 AAGTTCCTCTGAAACCAATCGTTACTGCTCTTACAGTGGTTCCTGGATGAGAAATGTC 2369

QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProThrAsp 840  
 Db 2370 CATTTGGACATACACAGATATTTACTAGATTTTATGAGAGGCTGGAAAGCATATAT 2429  
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyLysHisTyr 860  
 Db 2430 TTACAGATCTATCTCTCAGAGACAGACAGCATTAAGATGCTGAAATCGGAGAACTAT 2489  
 QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880  
 Db 2490 GAACCTGATCTTTTGCACTTCAAGAAACCTTGATCAGCATGATGCTCTTAAAA 2549  
 QY 881 ValIle 882  
 Db 2550 GTGATTA 2555  
 RESULT 13  
 AAD23843  
 ID AAD23843 standard, cDNA, 2510 BP.  
 AC AAD23843;  
 XX 07-MAR-2002 (first entry)  
 DT  
 XX  
 DE Human protease PRS-2 CDNA.  
 XX Human, protease, PRS-2; tranquilizer, gene therapy; vaccine; allergy;  
 KW infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
 KW atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
 KW gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
 KW epithelial disorder; uterine; anorexia; trauma; asthma; eczema; nausea;  
 KW hypertension; neurological disorder; Parkinson's disease; drug screening;  
 KW cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;  
 KW diabetes mellitus; glomerulonephritis; cardiovascular disorder; anemia;  
 KW autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
 KW developmental disorder; reproductive disorder; infertility; diarrhoea;  
 KW dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; sr.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 616..2358  
 FT /\*tag= a  
 FT /product= "Human protease PRS-2 protein"  
 FT  
 XX MO200183775-A2.  
 XX  
 PD 08-NOV-2001.  
 XX  
 PP 04-MAY-2001, 2001WO-US14651.  
 XX  
 XX 04-MAY-2000, 2000US-202082P.  
 PR 11-MAY-2000, 2000US-203566P.  
 PR 17-MAY-2000, 2000US-205803P.  
 PR 25-MAY-2000, 2000US-207477P.  
 PR 01-JUN-2000, 2000US-209402P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Deleage AM, Lal P, Hatalla A, Patterson C, Walla NK, Kearney L;  
 PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azinot Y, Elliott VS;  
 PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM;  
 PI Reddy R, Yue H, Tang YT;  
 DR WPI, 2002-034518/04.  
 DR P-PSDB, AAE14337.  
 XX  
 PT Novel human proteases and polynucleotides encoding the proteases,  
 PT useful for treating, diagnosing or preventing cell proliferative,  
 PT cardiovascular, autoimmune/inflammatory, neurological and developmental  
 PT disorders -  
 XX  
 PS Claim 5; Page 139-140; 151pp; English.

XX The invention relates to human proteases (PRTS-14) and its corresponding  
 CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for  
 CC the diagnosis, treatment and prevention of disorders associated with  
 CC increased or decreased expression of PRTS. Examples of such disorders  
 CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
 CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders  
 CC (AIDS, Addison's disease, allergy, anaemia, aschma, atopic dermatitis,  
 CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
 CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
 CC viral, bacterial, fungal, parasitic, protozoal and helminthic  
 CC infections); cardiovascular disorders (myocardial infarction, ischaemic  
 CC heart disease and hypertension); neurological disorders (epilepsy,  
 CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,  
 CC Parkinson's disease, stroke, mental disorders including mood, anxiety  
 CC and seasonal affective disorder and prion diseases); gastrointestinal  
 CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);  
 CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,  
 CC alopecia, scabies, insect bites and urticaria); reproductive disorder  
 CC (infertility, disruption of estrous and menstrual cycle and  
 CC gynaecomastia); and developmental disorders (renal tubular acidosis,  
 CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
 CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic  
 CC fragments are useful for screening libraries of compounds in several drug  
 CC screening assays. The present sequence is human protease PRTS-2 cDNA.

XX Sequence 2510 BP; 777 A; 494 C; 527 G; 712 T; 0 other;

# Alignment Scores:

Pred. No.: 0 Length: 2510  
 Score: 3970.50 Matches: 764  
 Percent Similarity: 89.81% Conservative: 3  
 Best Local Similarity: 89.46% Mismatches: 5  
 Query Match: 84.48% Indels: 82  
 DB: 24 Gaps: 6

US-10-070-464-1 (1-882) x AMD23843 (1-2510)

QY 42 TTPSERTGLNLEUYSYLVSEULEUAAASPTHRYGLYTRHISGLYTRMECETALA 61  
 DB 3 TGGAGTACGCTTAAAGAGCTGCTTGGCCGATCCAGAAATATCATGCTACATATGAGCT 62  
 QY 62 LYSALAPROHISAPHEMETPHEVALLYSARGASAPPROASGILYPROHISERASP 81  
 DB 63 AAGGACACATGATTTTCAITGTTGTGAAGAGATGATCCAGATGAGCTTCATTTCAGAC 122  
 QY 82 ARGILETYRLEUAAIAMECTSERGLYGLUASNAARGIUNSNTHRLEUPHETYSERGLU 101  
 DB 123 AGATCATTAACCTTGCATGCTGCGAGAGAACAGAGAAATACACGTTTTTATTCGAA 182  
 QY 102 ILEPROLYSTHRILEASNAAGALALAVALLUEUCLEUSERTPLYSPROLEUENASP 121  
 DB 183 ATTCGCCAAACTATCAATAGAGCAGCTTAAATGCTCTCTTGGAAAGCCCTTTTGGAT 242  
 QY 122 LEUPHEGLNALATHRIEUAAPRYGLYMETTYRSEARGLUGIUGIULEUENARGLU 141  
 DB 243 CTTTTCAGGCAACACTGAGCTAGATGAAATGTAATCTTGAGAGAGAACTATTAAAGAA 302  
 QY 142 ARGLYSARGILEGLYTHRVALLIYIIEALASERTYRAPHYRHHISGLINGISERGLYTHR 161  
 DB 303 AGAAAGCATTTGAAACAGTCGGAAATTCCTTCAATATATCCAAAGGAAGTGAACA 362  
 QY 162 PHELEUPHEGLNALAGLYSERGLYILETYRHHISVALYASAPGLYGLYPROGLINGLYPHE 181  
 DB 363 TTTCTGTTTCMAACCGGTAGTGGAAATTTATCACGTAAAGATGAGGGCCACAAAGGATTT 422  
 QY 182 THRGINGINPROLEUARGPROASILEUVALGUTHRSERCYSPROASNIIEATGMEASAP 201  
 DB 423 ACCCAACACCTTTAAGGCCCATCTAGTGAACCTAGTTGTTCCCAACATAAGGATGGAT 482  
 QY 202 PROLYSEUCYSPROALASAPPROASPTTPIIEALAPHEIIEHISSEASAPPILETTP 221  
 DB 483 CCAAAATATATGCTGCTGATCCAGACTGGATTCCTTTATATACATACCAAGATATTTGG 542

QY 222 ILESERASNIIEVALTHRARGGLUGIARGLEUTHRYRVALLHISASNIULEUAA 241  
 DB 543 ATATCTAACATCGTAAACAGAGAAAGAAAGAGACTCATTTATGTGCACAAATGAGCTACGC 602  
 QY 242 ASMETGLUGIWAAPALARGSERALGLYVALIATTHRPREVALLEUGIUGIUGIUGIUGI 261  
 DB 603 AACATGGAAGAAAGATGCGACATGCTGAGTCCGTACCTTTGTTCTCAAGAAATAATTT 662  
 QY 262 ASPARGTYRSEGLYTYRTPTRPCYAPROLYSALAGIUTRTHRPROSEGLYGLIYLS 281  
 DB 663 GATGATATTTTGGCTATTTGGTGTGTGTCCTCAAAAGCTGAACCACTCCAGTGTGTAA 722  
 QY 282 ILEUENARGILEUSTRYRGLUGIUNASAPGLUSERGIUVALGIUIEILEHISVALITHR 301  
 DB 723 ATTCCTGAAATTCATATGAAAGAAATGATGAAATCTGAGTGAATAATTATCATGTACA 782  
 QY 302 SERPROMETLEUGIUTRTHARG-ARGALIASPSETPHEARGTYRPROLYSTHNGLYTHRAL 321  
 DB 783 TCCCTATGTTGGAAACAAAGGCGAGATTCATTCGTTATCTTAAACAGGTACAGC 842  
 QY 321 AASPNULYSEVALTHRPHLYSMETSERGLUIIEMETILEASPALAGIYARGILEI 341  
 DB 843 AAATCCTRAAAGTCACTTTTAAGATGTCAGAAATATATGATGATCTGAGAGAGATCAT 902  
 QY 341 EASPVALLIIEAPLYSGIULEUILEGINPROPHEGIUIILEUPHEGIUVALGIUTY 361  
 DB 903 AGATGATATGATAGAAAGAACTAATTCAACCTTTGAGATTCATATTGAAAGAGTTGAATA 962  
 QY 361 RILEALARGALAGIYTRPTRPARGIUGIUGIUGIUGIUGIUGIUGIUGIUGIUGIUGI 381  
 DB 963 TATTCGAGAGCTGAGTGAAGTCTCTGAGGAAATATGCTTGTGTCATCTCACTAGATCG 1022  
 QY 381 GSERGLTHRARGLEUGIUILEVALLEUILESETPROGIUNUPHEIIEPROVALGIUAS 401  
 DB 1023 CTCCACAGACTCGCTTACAGATAGTGTGATCTCACCTGAATATTATTCAGATGAGAA 1082  
 QY 401 PASPVALLMETGLUARGIUNARGLEUILEGISERVALPROASPVETALTHRPROLEUI 421  
 DB 1083 TGATGTTATGGAAGAGCAGAGACTCATTTGAGTCAAGTCTGATCTGTGAGCCCATAT 1142  
 QY 421 EILETYRGLUGIUNTHRTHRAPPILETTPILEASNIIEHISAPPIIEPHEHISVALPHEPR 441  
 DB 1143 TATCTATGAAGAAACAAACAGACATCTGATTAATAATCAATGACATCTTCAATGTTTTTCC 1202  
 QY 441 OGLINSEHISGLUGIUGIUILEGIUPHEIIEPHEALSERGLUCYSYSTHRGIYPHEAR 461  
 DB 1203 CCAAAAGTCAAGAAAGAAATTTGATTTATTTTTCCTGAAAGCCAAACAGGTTTCG 1262  
 QY 461 GHISLEUTYRLYSILETHSERILEUUYSGIUSERLYSTYRYSARGSERSEGLYGI 481  
 DB 1263 TCATTATACAAATATCATCTATTATTTAAAGAAAGCAATATATACATCAAGTCAAGTGG 1322  
 QY 481 YLEUPROALAPROSERASPHELYSCYSPROIIIELYSGIUGIUIEALAIETHRSEGL 501  
 DB 1323 GCTGCTGCTCCAACT-----GTCACCT----- 1344  
 QY 501 YGIUTRPGIULVAL-----LEUGLY-----ARGHI 509  
 DB 1345 -----TGAMGATCATCATTCATGAGATCTTGAAGAACTCCATCTGTATGTGTGACACA 1400  
 QY 509 SGLYSERASNIIEGLIUNALAPGLIUNALARGATGLEUVALTYRPHEGIUGIYTHRYLVSAS 529  
 DB 1401 TATAGTTGAGATCCAGTGTATGATGAAGCAGAGAGCTGTATATTTGAAGCACAAAGA 1460  
 QY 529 PSEERPROLEUGIUIEHIISTEUTYRVALLISERTYRVALLASNPARGIUGIULVALITHRAR 549  
 DB 1461 CTCCTCTTAAGACATCACTGTACGTACGTATTAAGTAATCTTGAAGAGGTGACAG 1520  
 QY 549 GLEUTHRASPARGLIYTYRSEHISSERCYSGIIESEGLIHHISCYASAPPHPEHEI 569  
 DB 1529 GCTGATGACCGGTGCTACTACATTTCTGTCATCAGTCAAGCACTGTACTTTTAT 1580

QY 569 eSerIySTySerAnGlnLyAsnProHieCySvAlSerLeuTyLyLeuSerSerPr 589  
DB 1581 AAGTAAGATAGTAAACCAAGAAATCCACACTGTCCTTTCACCAAGTACAGTCC 1640  
QY 589 OGUAAPAPProThrCysLySThLySGluPheTrpAlaThrIleLeuAspSerAlaGI 609  
DB 1641 TGAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCCACCACTTTTGGATTCGACAGG 1700  
QY 609 YProLeuProAspTyThrProProGlnIlePheSerPheGluSerThrThrGlyPheTh 629  
DB 1701 TCCCTTCCTCACTATACCTCCACGAAATTTCTCTTTTGAAGTACTACTGCAATTAC 1760  
QY 629 rLeuTyGlyMetLeuTyLySTyProHieAspLeuGlnProGlyLySTySTyProThVa 649  
DB 1761 ATTGATTCGATGCTCTCAAGACCTCAATCATCAAGCTGGAAGAAATATCTACTCT 1820  
QY 649 lLeuPheIleTyGlyGlyProGlnValGlnLeuValAsnAsnArgPheLySGlyValLy 669  
DB 1821 GCTGTTCTATATATGCTGCTCCACAGTGCAGTGGTGAATATCGCTTTAAAGAGTCAA 1880  
QY 669 sTyPheArgLeuAnThrLeuAlaSerLeuGlyTyTrpAlaValValIleAspAsnAr 689  
DB 1881 GTATTTCCGCTTGAAATACCTCAAGCTCTCAAGTATATGTTGATGATAGACAAACAG 1940  
QY 689 gGlySerCysHisArgGlyLeuLySTyPheGlyValAlaPheLySTyLyMetGlyGlnI 709  
DB 1941 GCGATCCGTCACCGAGGCGCTTAATTTGAAGGGCCCTTAAATATAAAG----- 1992  
QY 709 eGluIlleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspPheIleAs 729  
DB 1992 ----- 1992  
QY 729 pLeuAspArgValGlyIleHisGlyTrpSerTyGlyGlyTyLeuSerLeuMetAlaLe 749  
DB 1992 ----- 1992  
QY 749 uMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpI 769  
DB 1993 -----GTTGCTATTGCTGGGCGCCACAGTCACTGCTGGAT 2027  
QY 769 ePheTyAspThrGlyTyTrpThrGluArgTyMetGlyHisProAspGlnAsnGluGln 789  
DB 2028 CTTCATATGATACAGATACAGAAACGTTATATGGGTACACCTTACCAAGTGAACAGGG 2087  
QY 789 YTyTrpLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArgLe 809  
DB 2088 CTATTACTTAGAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCTCTGAACCAAAATCGTT 2147  
QY 809 uLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLe 829  
DB 2148 ACTGCTCTTACATGCTTCCCTGATGAGAATGTCATTTTGCACATACCACTATATTAAT 2207  
QY 829 uSerPheLeuValArgAlaGlyLySTyProTyAspLeuGlnIleTyProGlnGluArgHis 849  
DB 2208 GAGTTTCTTACTGAGGCTGGAAGCCATATGATTTA-----CAGGAGACACA 2255  
QY 849 sSerIleArgValProGluSerGlyGluHisTyGlyLeuHisLeuLeuHisTyLeuGly 869  
DB 2256 CAGCATTAAGAGTTCCTGAATCGGAGAAACATTAATGAATGCTATTTTGGACCTACCTTCA 2315  
QY 869 nGluAsnLeuGlySerArgIleAlaIleAlaLeuLyValIle 882  
DB 2316 AGAAAACTTGATCAGTATCTGCTGCTTAAAGTATA 2355

## RESULT 14

ABNS9775 standard; cDNA; 2668 BP.

ABNS9775;

28-JUN-2002 (first entry)

Novel human coding sequence SEQ ID NO: 186.

XX Human; antihaemic; vulnery; antiinflammatory; immunomodulator;  
KM antinfertility; cerebroprotective; cyostatic; rheumatic; gene therapy;  
KM neuroprotective; antiparkinsonian; protein therapy; EST;  
KM expressed sequence tag; gene; ss.  
OS Homo sapiens.  
XX WO200222660-A2.  
XX 21-MAR-2002.  
XX 10-SEP-2001; 2001WO-US26015.  
XX 11-SEP-2000; 2000US-0659671.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX WPI; 2002-292408/33.  
XX P-PSDB; ABB97362.  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -  
XX Claim 1; SEQ ID NO 186; 509pp; English.  
XX The present invention provides the protein and coding sequences of 444  
CC novel hman proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a coding sequence of the  
CC invention.  
XX SQ Sequence 2668 BP; 796 A; 564 C; 592 G; 716 T; 0 other;  
Alignment Scores:  
Score: 0 Length: 2668  
Percent Similarity: 3771.00 Matches: 724  
Best Local Similarity: 82.09% Conservative: 0  
Query Match: 80.23% Indels: 158  
DB: 24 Gaps: 2  
US-10-070-464-1 (1-882) x ABNS9775 (1-2668)  
QY 1 MetAlaAlaIleMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20  
DB 234 ATGCGACACCAATGAGAAACGAACAGCTGGGTGTTAGATATTGAAATCGCGAACTGT 293  
QY 21 GluGluAsnIleGluSerGlnAspArgProLyLeuGluProPheTyValGluArgTy 40  
DB 294 GAGAGAAATATTAATACACAGATCGGCTTAAATTTGAGACCTTTATGTTGACGGTAT 353  
QY 41 SerTrpSerIleuLySTyLeuLeuAlaAspThrArgTyTrpHisGlyTyTrpMetMet 60  
DB 354 TCCCTGAGTCAAGCTTAAAGAGCTGCTGCGATACCAAGAAATATCTGCTCATGATG 413  
QY 61 AlAluAlaProHisAspPheMetPheValLyArgAsnAspProAspGlyProHisSer 80  
DB 414 GCTAAGCACACATGATTTTCATGTTTGTGAAGAGAAATATCAAGATGACCTCATTTCA 473  
QY 81 AspArgIleTyTrpLeuAlaMetSerGlyGluAsnArgGluAnThrLeuPheTySer 100  
DB 474 GACAGATCTATTAATCTTGCATGCTGGTGAAGAACAGAAATAACACTGTTTATTC 533

QY 101 GluIleProLysThrIleAsnArgAlaIleValLeuMetLeuSerTrpLysProLeu 120  
DB 534 GAATTCCTCCAAACTATCATAGACAGCAAGCTTTAAGCTCTCTTGAAGCCTTTTG 593  
QY 121 AspheuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140  
DB 594 GATCTTTTTCAG----- 605  
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160  
DB 605 ----- 605  
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180  
DB 605 ----- 605  
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
DB 606 -----CAACAACCTTTAAGGCCCAATCTAGTGAATCTAGTTGTCACATACGAGATG 659  
QY 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220  
DB 660 GATCCAAATTAATATGCGCTGATCCAGACTGGAATTCCTTTATACATAGCAACGATATT 719  
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240  
DB 720 TGGATATTAACTGCTAAGACAGAAAGAAAGAAAGAACTCACTTATGTCACAAAGAGCTA 779  
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260  
DB 780 GCCAAACATGAAGAAGATGCCAGATCAGCTGGAAGTCGCTACCTTTGTTCTCCAAAGAA 839  
QY 261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280  
DB 840 TTTTATGATATTTCTGCTATTTGGTGTGCCAAAGCTGAACAACTCCACAGTGTGTGT 899  
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300  
DB 900 AAAATTCTTAAATTTCTATAGAACAAATATGATATGAGTGGAAATTTATTCATTT 959  
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
DB 960 ACATCCCTATGTTGGAAACAAGAGGCGAGATTCATTCGGTTATCTTAAACAGGTGCA 1019  
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340  
DB 1020 GCAATCTCTAAAGTCATTTAAGATGTCAGAAATATGATGATGCTGAAGGAAAGGATC 1079  
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGlnIleLeuPheGluGlyValGlu 360  
DB 1080 ATGATGTCATACATAGAAAGAACTAATTCACCTTTTGAGATTTCTAATTGAAGAGGTGAA 1139  
QY 361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380  
DB 1140 TATATTGGCAGAGCTGAGTGAATCTCTGAGGAAATATGCTTGATTCATCTCTAGAT 1199  
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
DB 1200 CGCTCCCAAGACTCGCTCAACATAGTGTGATCTCACCTGAATTAATTATCCAGTGA 1259  
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420  
DB 1260 GATATGTTATGAAAGGCAAGACTCATTTAGTCAGAGCTGATTCGTGACGCCACTA 1319  
QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440  
DB 1320 ATTATTCATGAAGAAACAACAGACATCTGATTAATATCCATGACATCTTTCATGTTT 1379  
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460  
DB 1380 CCCCAAGATCAGAAAGGAAATGAGTTAATTTTGTCTCGAATGCAAAACAGGTTTC 1439  
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480

DB 1440 CGTCATTTATACAAATTAATACATCTATTTTAAAGAAACAAATATTAACGATCCAGTGT 1499  
QY 481 GlyLeuProAlaProSerAspPheLysCysAspProIleLysGluGluIleAlaIleThrSer 500  
DB 1509 GGGCTGCTGCTCCCAAGATGATTCAGAGTCTCTATCAAGAGAGAGATGACATTCACAGT 1559  
QY 501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520  
DB 1560 GGTGAATGGAAAGTTCTTGGCGGCGATGATCTTAATATCCAAAGTTGATGAATCGAAAG 1619  
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGlnIleHisLeuTyrValIleSer 540  
DB 1620 CTGGTATATTTTGAAGGCAACCAAGACTCCCTTTAGAGCATCAGCTTACGTAGTCACT 1679  
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560  
DB 1680 TACGTAATCTCGAGAGAGTGAACAAGGTGACTGACCGTGGCTACTCATCTTGTGCTGC 1739  
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580  
DB 1740 ATCAGTCAGCACTGTGACTTCTTTATAGATAGATAGTAACCAAGAAATCCACACTGT 1799  
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600  
DB 1800 GTGTCCCTTACAAAGCTATCAAGTCTGAAAGATGACCCAACTTGCAAAACAAAGAAATTT 1859  
QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe 620  
DB 1860 TGGGCACACATTTTGATTCAGCAGGTCTCTTCCTGACTATACCTCCCAAGAAATTTTC 1919  
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640  
DB 1920 TCTTTTGAAGTACTACTGATTTTACATTTGATGGAGGCTCTCAACCCCATATATCTA 1979  
QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660  
DB 1980 CAGCTGGAAGAAATATCTTACTGTCTGTTCATATATGGTGGTCTCAG----- 2030  
QY 661 ValAsnAspArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680  
DB 2030 ----- 2030  
QY 681 TyrValValValValIleAspAsnArgGlySerCysHisAlaGlyLeuLysPheGluGly 700  
DB 2030 ----- 2030  
QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720  
DB 2030 ----- 2030  
QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleGlyTyrSerTyr 740  
DB 2030 ----- 2030  
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760  
DB 2031 -----GTTGCTGATTT 2039  
QY 761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspTrpGlyTyrThrGluArgTyrMet 780  
DB 2040 GCTGGGCGCCCAAGTCACTCTGAGATCTTCTATGATACAGATACACCGAAACGTTATAG 2099  
QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800  
DB 2100 GGTCAACCTGACCAAGATGAAACAGGGCTATTAAGATCTGTGGCCATGACAGCAADA 2159  
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820  
DB 2160 AAGTTCCCTCTGAAACCAAAATCGTTTACTGCTTACATGGTCTTCGGATGAGAAATGTC 2219  
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyAlaGlyLysProTyrAsp 840

DB 2220 CATTTCGACATACCATATATTACTGAGTTTTCAGTGGCGTGAAGGCATATGAT 2279  
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlnSerGlyGlnHisTyr 860  
 DB 2280 TTACGATCTATCTCTCAGAGAGACAGCATPAGATGCTCGAATCGGAGAACATTAT 2339  
 QY 861 GluLeuHisIleLeuHisTyrIleGlnGluLeuGlnArgIleValIleValIleLeuHis 880  
 DB 2340 GAACGTCATCTTTGACATCACTTCAGAAAACCTTGATCCGATATGCTGCTAATAA 2399  
 QY 881 ValIle 882  
 DB 2400 GTGATA 2405  
 RESULT 15  
 ABK83328  
 ID ABK83328 standard; cDNA; 4309 BP.  
 AC ABK83328;  
 DT 12-AUG-2002 (first entry)  
 DE cDNA encoding human DPRP-1 splice variant #4.  
 XX  
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;  
 KW DPRP; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KW dyskinesia; reproductive disorder; inflammatory disorder;  
 KW metabolic disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200231134-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PE 12-OCT-2001; 2001WO-US31874.  
 XX  
 PR 12-OCT-2000; 2000US-240117P.  
 XX  
 PA (FERR ) FERRING BV.  
 XX  
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;  
 XX  
 DR WPI; 2002-444178/47.  
 DR P-PSDB; ABG61597.  
 XX  
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
 PT viral infections, cancers, allergies, neurological disorders, or pain  
 PT  
 PS -  
 XX  
 PS Disclosure; Page 67-68; 113pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.  
 CC  
 SQ Sequence 4309 BP; 1304 A; 790 C; 907 G; 1308 T; 0 other;

Alignment Scores:  
 Pred. No.: 0 Length: 4309  
 Score: 3661.50 Matches: 708  
 Percent Similarity: 80.09% Conservation: 0  
 Best Local Similarity: 80.09% Mismatches: 1  
 Query Match: 77.90% Indels: 176  
 DB: 24 Gaps: 1  
 US-10-070-464-1 (1-882) x ABK83328 (1-4309)  
 QY 1 MetAlaAlaIleMetGluThrGlnGlnLeuGlnValGluIlePheGluThrAlaAspCys 20  
 DB 214 ATGGCAGCAGCATGTAAGAAACAGAACAGCTGGGTGTGAGATATTGGAACCTCGACATGT 273  
 QY 21 GlnGluAsnIleGlnSerGlnAspArgProIleGlnLeuGlnProPheTyrValGlnArgTyr 40  
 DB 274 GAGGAGAAATATGAAATCAGAGATGCGCTAAATATGAGCCCTTTTATGTTAGCGGTAT 333  
 QY 41 SerTPSerGlnLeuValGlnLeuValAlaSerThrArgValGlnTyrHisGlyTyrMetMet 60  
 DB 334 TCTGGAGTCAGCTTAAAGACCTGCTCCATACCAAGAAATATCATGGCTAACATGATG 393  
 QY 61 AlaValAlaProHisAspPheMetPheValIleValArgAsnAspProAspGlyProHisSer 80  
 DB 394 GCTAAGCAGCAGCATGATTTTCATGTTTGTGAAGAGAAATGATCCAGATGACCTCATTTCA 453  
 QY 81 AsparGlnIleTyrTyrLeuAlaMetSerGlyGlnAsnArgGlnAsnThrLeuPheTyrSer 100  
 DB 454 GACAGAAATCTATTACCTTGCCATGCTGTGAGAAACAGAAATAACACTGTTTATTCT 513  
 QY 101 GluIleProIleValIleAsnArgAlaValIleMetLeuSerThrIleProIleLeu 120  
 DB 514 GAAATTTCCCAAACTATCAATAGAGACAGCTTAAATGCTCTTGGAAAGCTCTTTTG 573  
 QY 121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGlnGluGluLeuLeuArg 140  
 DB 574 GATCTTTTCAGGCAACCTGACATGGAATGATTTCTCAGAAAGAAATATTATTAAGA 633  
 QY 141 GluArgGlyValGlnIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnIleSerGly 160  
 DB 634 GAAAGAAAACCATTTGGAACAGCTCGGAATTCCTTTCATGATTAACCAAGAAAGTGA 693  
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIleValAspGlyGlnProGlnGly 180  
 DB 694 ACATTTCTGTTCAAGCCGTAAGTGAATTTATATCAGTAAAGATGAGAGGCGCACAGAGA 753  
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
 DB 754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGAATAGTGTGCCAACATACGATG 813  
 QY 201 AspProIleValCysProAlaAspProAspThrIleAlaPheIleHisSerAsnAspIle 220  
 DB 814 GATCCAAAATTAATGCGCTGATCCAGATCGATGCTTTTATATCATATGCAACATATAT 873  
 QY 221 TrpIleSerAsnIleValThrArgGlnGluArgValGlnLeuThrTyrValHisAsnGluLeu 240  
 DB 874 TGATATATTAACTGTAACCAAGAAAGAAAGACCTCATGATGTCACATAGCTA 933  
 QY 241 AlaAsnMetGlnGluAspAlaArgSerAlaGlyValAlaThrPheValIleGlnGlnGlu 260  
 DB 934 GCCAACATGAAAGAAAGATGCCAGATCAGTGAAGTCCGCTTCTTGTCCAAAGAA 993  
 QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProIleValGluThrThrProSerGlyGly 280  
 DB 994 TTTGATATATATTCTGGCTATGTGTGTGTCCAAAAGCTGAACCAACCTCCAGTGTGT 1053  
 QY 281 LysIleLeuArgGlnLeuTyrGlnGluAsnAspGlnSerGlnValGluIleIleHisVal 300  
 DB 1054 AAAATTTCTTAATTTCTATGAAAGAAATATGAAATTTGAGGTGGAATTTATTTAT 1113  
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProIleThrGlyThr 320  
 DB 1114 ACATCCCTATGTGGAAACAGAGAGGCGAGATTCATTCCTGTTATCTTAAACAGGTACA 1173

```

Oy 321 AlaasnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db 1174 GCAAAATCCTAAAGTACACTTTTAAAGTGTCAAGAAATAGATTGATGCTGAAGAAAGATC- 1232
Oy 341 IleasPValIleasPLeuLeuIleGlnPropheGluIleLeuPheGluGlyValGlu 360
Db 1232 -----
Oy 361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
Db 1232 -----
Oy 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db 1232 -----
Oy 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db 1232 -----
Oy 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db 1232 -----
Oy 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThrGlyPhe 460
Db 1232 -----
Oy 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db 1232 -----
Oy 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db 1232 -----
Oy 501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db 1233 -----CAAGTTGATGAAGTCAAGAG
Oy 521 LeuValIleTyrPheGluGlyThrLysAspSerProLeuGluHisIleLysLeuTyrValIleSer 540
Db 1254 CTGGTATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGTACGTAGTCACTG 1313
Oy 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCys 560
Db 1314 TACGTAAATCCTGAGAGGTACAAAGCTGACTGACCTGCTACTCACAATTCTTGCTGC 1373
Oy 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db 1374 ATCAGTCAAGCACTGTGACTTTTATTAAGTATAGTAAACCAAGAAATCCACACTGT 1433
Oy 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db 1434 GTGCCCTTTAACAAGCTATCAAGTCTGAAGATGACCAACTGCAAAACAAGAAATTT 1493
Oy 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db 1494 TGGGCCACCAATTTGGATTCGACAGGTCCTCTTCTGACTATACTCTCCGAAATTTTC 1553
Oy 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db 1554 TCCTTTGAAAGTACTACTGATTTACATTTGATGGATGCTCTCAAGCCCTCATGATCTA 1613
Oy 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnL 660
Db 1614 CAGCCTGGAAGAAATATCTACTGTCTGTTCATATATGCTGTCTCTCAGCGTCAGT 1673
Oy 660 euValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeu 680
Db 1674 TGGTGAATTAATCGATTAAAGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCAG 1733

```

```

Oy 680 IYTYRValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluG 700
Db 1734 GTTATGCTGCTGTATGATGTATACAAACAGGGATCTCTGACAGGCGCTTAAATTTGAAG 1793
Oy 700 IYAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnT 720
Db 1794 GCGCCTTAAATATTAATAATGGGTCAAAATTAAGATTAAGATCAAGGTGAAGGACTCCAAAT 1853
Oy 720 YRLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerT 740
Db 1854 ATCTAGCTTCTCGATATGATATTTCACTTACATTCGATGAGGACATCCACGCGTGGTCT 1913
Oy 740 YRGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI 760
Db 1914 ATGAGGATACCTCTCCCTGATGGCATTAATGCAAGGATCAAGATTTCTTCAGGGTTGCTA 1973
Oy 760 IeAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrM 780
Db 1974 TTGCTGGGGCCCAAGTCACTGTGATCTTCTATGATACAGGATACACGGAAAGCTATTA 2033
Oy 780 eTGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaG 800
Db 2034 TGGGTCAACCCGTACAGCAAGATGACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAG 2093
Oy 800 IuLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnV 820
Db 2094 AAAAGTTCCTCTGAAACCAATCGTTTACTGCTTTACATGATGTTTCTTGGATGGAAATG 2153
Oy 820 aHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrA 840
Db 2154 TCCATTTTGCATACACAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAGCATATG 2213
Oy 840 ePLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisT 860
Db 2214 ATTTACAGATCTATCTTCAGAGAGACACAGCATTAAGTTCCGAAATCCGGAGAACAT 2273
Oy 860 YRGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuL 880
Db 2274 ATGAACTGCATCTTTTGCATCACTTCAAGAAACCTTGATACGATATTCCTGCTTAA 2333
Oy 880 YSValIle 882
Db 2334 AAGTGATA 2341

```

Search completed: October 15, 2003, 17:56:50  
 Job time : 780.054 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:29:02 ; Search time 138.181 Seconds

(without alignments)  
2817.309 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAAMETQLGVEIFETAD.....HLHLYGELSGRIALAKVI 882

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 segs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODEL=frame\_plus\_model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10070464/runat\_15102003\_113554\_24845/app\_query.fasta\_1.26552  
-DB=Issued\_Patents\_NA -QEXT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi  
-LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pro -NORF=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10070464@cgn2\_1\_1146@runat\_15102003\_113554\_24845 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	755.5	16.1	823	US-09-280-116-171	Sequence 171, App
2	597.5	12.7	5496	US-09-462-284-1	Sequence 1, Appl
3	529	11.3	3407	US-10-002-593-5	Sequence 1, Appl
4	525	11.2	2924	PCR-US93-07923-1	Sequence 1, Appl
5	454	9.7	2815	US-08-230-491A-1	Sequence 1, Appl
6	454	9.7	2815	US-08-619-280A-1	Sequence 1, Appl
7	454	9.7	2815	US-08-940-391-1	Sequence 1, Appl
8	404	8.6	612	US-09-392-184-31	Sequence 31, Appl
9	339.5	7.2	4982	US-08-699-103B-1	Sequence 1, Appl
10	339.5	7.2	4982	US-09-223-059-1	Sequence 1, Appl
11	339.5	7.2	4982	US-09-628-133-1	Sequence 1, Appl
12	328.5	7.0	657	US-09-221-017B-646	Sequence 646, App

13	328	7.0	543	US-09-221-017B-253	Sequence 253, App
14	282	6.0	1896	US-09-016-080-2	Sequence 2, App
15	278.5	5.9	535	US-09-389-681-428	Sequence 428, App
16	278.5	5.9	535	US-09-620-405B-428	Sequence 428, App
17	278.5	5.9	535	US-09-433-826B-428	Sequence 428, App
18	278.5	5.9	535	US-09-604-287A-428	Sequence 428, App
19	232	4.9	3085	US-09-221-017B-1010	Sequence 1010, App
20	193	4.1	502	US-09-280-116-172	Sequence 172, App
21	182	3.9	1869	US-08-664-646A-1	Sequence 1, Appl
22	182	3.9	1869	US-09-066-285-1	Sequence 1, Appl
23	182	3.9	1869	US-09-261-006-1	Sequence 1, Appl
24	182	3.9	1869	US-08-951-088-1	Sequence 1, Appl
25	182	3.9	1869	US-09-609-556-1	Sequence 1, Appl
26	182	3.9	1869	US-09-609-570-1	Sequence 1, Appl
27	182	3.9	1869	US-09-427-372-1	Sequence 1, Appl
28	182	3.9	1869	US-09-693-554-1	Sequence 1, Appl
29	175.5	3.7	1845	US-09-252-991A-3489	Sequence 3489, App
30	157.5	3.4	815	US-09-221-017B-962	Sequence 962, App
31	133	2.8	1032	US-09-724-623-17	Sequence 17, Appl
32	133	2.8	1230025	US-09-198-452A-1	Sequence 1, Appl
33	128.5	2.7	904	US-09-221-017B-58	Sequence 58, Appl
34	128	2.7	536165	US-09-214-808-1	Sequence 1, Appl
35	125.5	2.7	5837	US-07-686-340-1	Sequence 1, Appl
36	125.5	2.7	5837	US-08-004-139B-1	Sequence 1, Appl
37	125.5	2.7	5837	US-08-117-491-1	Sequence 1, Appl
38	125.5	2.7	5837	US-08-271-364A-1	Sequence 1, Appl
39	125.5	2.7	5837	US-08-811-492-1	Sequence 1, Appl
40	125.5	2.7	5837	US-08-222-715B-1	Sequence 1, Appl
41	125.5	2.7	5837	PCR-US96-10545A-1	Sequence 1, Appl
42	124	2.6	2636	US-08-227-689-1	Sequence 1, Appl
43	124	2.6	2636	US-09-368-169-1	Sequence 1, Appl
44	118.5	2.5	4403765	US-09-103-840A-2	Sequence 2, Appl
45	117.5	2.5	4411529	US-09-103-840A-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-280-116-171  
Sequence 171, Application US/09280116A  
Patent No. 6331427  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs  
FILE REFERENCE: 5800-24, 035800/176965  
CURRENT APPLICATION NUMBER: US/09/280,116A  
CURRENT FILING DATE: 1999-03-26  
NUMBER OF SEQ ID NOS: 268  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 171  
LENGTH: 823  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: prolyl oligopeptidases  
US-09-280-116-171

#### Alignment Scores:

Pred. No.: 1.96e-77 Length: 823  
Score: 755.50 Matches: 164  
Percent Similarity: 88.95% Conservative: 5  
Best Local Similarity: 86.32% Mismatches: 16  
Query Match: 16.07% Indels: 16  
DB: 4 Gaps: 2

US-10-070-464-1 (1-882) x US-09-280-116-171 (1-823)

QY 708 GlnGluGluLeuAspAspGlnValGluGly-LeuGlnIlyrLeuAlaSerArgTyrAspPh 727  
DB 3 CAATAGAAATGTGACATGACATGACATGACATGACATGACATGACATGACATGACATG 62  
QY 727 eileAspLeuAspArgValGlyIleGlyTyrSerTyrGlyGlyTyrLeuSerLeuMe 747  
|||||

```

Db      63 CATTGACTTAGATCGTGGGATCCACGGCTGCTCTATGAGATACCTCTCCCTGAT 122
Qy      747 fAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThre 767
Db      123 GGCATTAAAGCAGAGGTGAGATATCTTCAGGGTTGCTATGCTGGGTCCCACTCATCT 182
Qy      767 uTTPilePheTyArgPThrGlyTyThrGluArgTyMetGlyHisProAspGlnAsnG 787
Db      183 GTGGATCTTCTATGATACAGATACAGAACCTTATATGGTCACTGACCCGACCAATGA 242
Qy      787 uGlnGlyTyTyLeuGlySerValAlaMetGlnAlaGluTyPheProSerGluProAs 807
Db      243 ACAGGGCTATTAATTAGATCTGTGGCCATGACAGCAAGAAAGTTCCCTGAAACAA 302
Qy      807 nArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSer 827
Db      303 TGGTTACTGCTCTTACAGAGGTTCTCGATGAGAAATGTCATTTTGACATACACGAT 362
Qy      827 eLeuLeuSerPheLeuValArgAlaGlyLysProTyArgPLeuGln---IleTy----- 844
Db      363 ATTACTGAGTCTTTTAGTAGAGGGCTGGAAGCAGATGATTTACGATATCTTATTTT 422
Qy      845 -----ProGlnGluArgHis-SerIleArgVal-ProGluSer-G 857
Db      423 GTTGTGTGATAGATCTATCTCAGGAGAGACACCGGCATTAAGAGTTCCCTGATCGGG 482
Qy      857 uGlnHisTyGlyLeuHisLeuLeu-HisTyLeuGlnGlu-AsnLeuLysSer-Arg 876
Db      483 AATAACCTTATGAACTGCATCTTTTGACACTTACCTTCAAGAAAACCTTGATCACTTAT 542
Qy      876 eAlaAlaLeuLysVal 881
Db      543 TGCTGCTCTTAAATT 558

RESULT 2
US-09-462-284-1
; Sequence 1, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Afifolter, Michael
; APPLICANT: Van den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; TITLE OF INVENTION: ASPERGILLUS ORYZAE
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5496
; TYPE: DNA
; ORGANISM: Fungus
US-09-462-284-1

Alignment Scores:
Pred. No.: 9 566-58 Length: 5496
Score: 597.50 Matches: 208
Percent Similarity: 42.67% Conservative: 112
Best Local Similarity: 27.73% Mismatches: 278
Query Match: 12.71% Indels: 152
DB: 4 Gaps: 33
US-10-070-464-1 (1-882) x US-09-462-284-1 (1-5496)
Qy      145 lIleglyThrValGlyIleAlaSerTyArgPThrHisGlnGlySerGlyThrPhe----- 162
Db      2239 GTCCGTGCTGTGGCGTCAACCA-----CACCAAGCAGTATCGGCATTCCTTCTT 2289
Qy      163 -----LeuPheGlnAlaGlySerGlyIleTyThrHisValLysAspGlyLysProGlnGly 180

```

```

Db      2290 TCCGATTAATTACGTCCAGAGTGTGAGTCACTCAAGTCCG----- 2331
Qy      181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db      2332 -----GCCCCTGATCCCGATCA 2349
Qy      201 AsPro-----LysLeuCysProAlaAsPro-----AspTrp-IleAlaPheIleHis 216
Db      2350 GGAAGGTGATATTATTCATATGATGCAATGAGCCCGCGCAATATGATCGCTTTGCTTGG 2409
Qy      216 sSerAsnAsp-----IleTrpIleSerAsnIleValThrArgGluGluArgLeuThr 234
Db      2410 CGAGATATGACCTTATATGCTGGGATATATGATCGCTTACTGCC----- 2452
Qy      234 rTyValHisAsnGluLeuAlaAsnMetGluLysAla-----ArgSerAlaG 251
Db      2453 -----ATTACTGATATGATGATGCGCCGACATGTCACCG 2487
Qy      251 yValAlaThrPheValLeuGlnGluGluPhe-----AspArgTySerGlyTyThrP 269
Db      2488 CGTGCCGAGCTGGATCTATGAAAGAGAGATCTCGGCATTCGCTACGCG--TTGGTGT 2544
Qy      269 pCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyGluG 289
Db      2545 CTCGCCAGATGGTGA-----TATCGCTTACTTGAG 2577
Qy      289 uAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThrArg 309
Db      2578 CTTCAATGAGACTGGGGTTCCGACCTACACCGTTCAGTATTTATGATTAACCAAGAGAT 2637
Qy      309 gAlaAspSerPhe-----ArgTyProLysThrGlyThrAlaAsp 323
Db      2638 CCGTCCGGGTATCCATGAGAGCTGAAATGATATCCCAAGGTGCGCAGACAGATTC 2697
Qy      323 oLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleAsp 343
Db      2698 GACCGTGAAGTGAAT-----CTCCTTAACAT 2724
Qy      343 lIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlyTyTrp 363
Db      2725 CGCTAGCAAGAGGTGAAGCAGCGCCGATGACCGCTTCGATCAACGATGATGATCAT 2784
Qy      363 aArgAlaGlyTrpThrProGluGlyLysTyAlaTrpSerIleLeuLeuAspArgSer 383
Db      2785 T-----GGGAGAGTGGCTTGG-----CTCAGTGAATCACTCACAC 2817
Qy      383 nThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGluAspAsp 403
Db      2818 CACC-----GTTGCTGCTAAGCGCTTCAACCGTGTCCAGAC----- 2854
Qy      403 lMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeuIleTy 423
Db      2855 -CAGCAAAAGTGTGCGGTGATATCTCGCAACCAAGCTACT---GTCATACAGCA 2910
Qy      423 rGluGluThrThrAspIleTrpIle---AsnIleHisAspIlePheHisValPheProG 442
Db      2911 CCGAGATGAGACCGATGATGATGCTGATACCTTTTTCATGAAGATATTTGGCCCTAT 2970
Qy      442 nSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPheArg 462
Db      2971 CAAGCGTCCGACAAAGATGCTCTACTATGACATCTCTGACCTTGGGATGGCGCA 3030
Qy      462 eLeuTyTrpLysIleThrSerIleLeuLysGluSerLysTyTrpLysArgSerSerGly 482
Db      3031 TCTGTATCTCTTCCC-----GTTTGGGGGGGCGCA 3060
Qy      482 uProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSerGly 502
Db      3061 ACCTATCCCA-----CTAACCAAGGCGCA 3084
Qy      502 uTTPGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArgLeuVal 522

```

```

Db      3085 CTGGAGGTCACG-----TCTATTCTGAGTATTGATCAGAAAGCCAGTTGGT 3132
Qy      522 1TyrPheGluGluGlyThrIlyAspSerProLeuGluHisIleuTyValValSerTyVa 542
Db      3133 GTACTACTGTCGTGACTCAACACCAAGCAGCCAGGCCATCTCTACTCCGTCTCTAT-- 3190
Qy      542 1AsnProGluGluValThrArgLeuThrAspArgGlyTyrSer-----HisSerCyeCy 560
Db      3191 -TCCACGTTGGGGTACCCCGCTGTCGACACACACCGTTGGCGGTACTGGTCTTC 3249
Qy      560 s1IleSerGlnHisCysAspPhePheIleSerIleTySerAsnGlnIlyAsnProHisCy 580
Db      3250 CTTCCTCCGGAACCTGGGCTACTACATCTCTACATACGAGGCGCAGACGTACCTAC-- 3307
Qy      580 sValSerLeuTyIlyIlyLeuSerSerProGluAspAspProThrCyIlyIlyGluIly 600
Db      3308 -CAGGAACCTTACAGCAGCCACAGT-----ACCAAACTACT 3342
Qy      600 eTrrAlaThrIleLeuAspSerAlaGlyProLeu-----ProAspTyThrProP 617
Db      3343 C---GGCAATACACCGACAGCCAAAGTACTCGAAGAAATCAAGAGTATGATGCG 3399
Qy      617 oGluIlePheSerPheGlu-----SerThrThrGlyPheThrLeuTyIlyMetLeuTy 635
Db      3400 CAACATCACTACTCTGAGCTTCCCTCCCTCCGAGAAACCTCAATGTATGATCAGCG 3459
Qy      635 IlyIserProHisAspLeuGlnProGluIlyIlyIlyTyProThrValLeuPheIleTyGly 655
Db      3460 CTTACCCCCCGGGTCTCCCGGATTAAGAAATCCCATATCTTTCACCCCATAGCGCG 3519
Qy      655 yProGlnValGlnLeuValAsnAsnAspPheIlyGlyValIlyTyIly---PheArgLeuAs 674
Db      3520 CCCAGCGCCGCAAGAGTACCAAGAGATGGCAGCGCCGAATTTCAGAGCCATATGTGCG 3579
Qy      674 nThrLeuAspSerLeuGlyTyIlyValValValIleAspAsnArgGlySerCysHisArg 694
Db      3580 CTCACAGCGCAACTCGAGTACGTAACTGACGTGCAACCCGCGACAGGTTTCAA 3639
Qy      694 gGlyLeuAspPheGluGluGlyAlaPheIlyTyIlyMetGlyGlnIleGluIleAspAsp 714
Db      3640 AGGAGCAAGTTCCTCCCTCCGCTACACCGCGCCGCTCCCTCGAAGCAAGAACCA 3699
Qy      714 nValGluGlyLeuGlnTyIlyLeuAlaSerArgTyIlyAspPheIleAspLeuAspArgVal 734
Db      3700 GATCTACCGCGCGCAACAG--GCGGCCAATCCCTCGATCGATGACAGACACATCGG 3756
Qy      734 yIleHisGlyTyIlyPheTyIlyGlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIly 754
Db      3757 CATCTGGGGCTGAGTTTCGGAGGCTACTTGAACCAAGCTCTCGAAGAACAGACAGCGG 3816
Qy      754 PilePheArgValAlaIleAlaGlyAlaProValThrIleTrrIlePheTyIlyAspThr 774
Db      3817 TGCCTTCAATTAGAGATCATACGCGCCCTGTCTTCTGACTGGCGTTCTTACAGTCAAT 3876
Qy      774 yTyThrGluArgTyIlyMetGlyHisProAspGlnAsnGluGlnGlyTyIlyTyIlyGly 794
Db      3877 GTACCGGAGGCTCATGATGAACCCCTCGACCAATGAGAGGCTAGAGACAGAGCGC 3936
Qy      794 rValAlaMetGlnAlaGluIlyPheProSerGluProAsnArgLeuLeuLeuHisGly 814
Db      3937 CGTCTC---CGCAAGACTGACGGGTTCAAGAACTCGAGGGCGGATCTTGATCTACACAG 3993
Qy      814 yPheLeuAspGlnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAr 834
Db      3994 AACGGGCAACATACGTCATTTCCAGAACTCGGCTGGCTGGTGATCTCTTAATGGG 4053
Qy      834 gAlaGlyIlyIlyProTyIlyAspLeuGlnIleTyIlyProGlnGluArgHisSerIleArgVal 854
Db      4054 CGATGCG-----GTCTCTCTGGAAGCTCATTTGCAATGGTTGAC 4095
Qy      854 oGluSerGlyIlyHisIlyTyIlyGluLeuHis 863
Db      4096 AGACTCAGACCAAGCAATCACTACCAT 4123

```

```

RESULT 3
US-10-002-593-5
; Sequence 5, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOGENSIN CONVERTI
; FILE REFERENCE: Acty Docket No. 6586198 1242/48/2
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US/10/002,593
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3407
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-593-5

Alignment Scores:
Pred. No.: 3,44e-50 Length: 3407
Score: 529.00 Matches: 186
Percent Similarity: 41.96% Conservative: 109
Best Local Similarity: 26.46% Mismatches: 280
Query Match: 11.26% Indels: 128
DB: 4 Gaps: 32

US-10-070-464-1 (1-882) x US-10-002-593-5 (1-3407)

Qy      206 ProAlaAspProAspTrrIle-----AlaPheIleHis 216
Db      520 CCMAACACACACAGTGGGTACATGTCACCAAGTGGGTCAATTAATGGCATATGTTGG 579
Qy      217 SerAsnAspIleTrrIleSerAsnIleValThrArgGluGluArgLeuThrTyVal 236
Db      580 AACCAATGACATTTATGTTAAATTAATGACCAAAATTTACCAAGTTTACCAATTCACATGGAGC 639
Qy      237 HisAsnGluLeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheVal 256
Db      640 GGG-----AAAGAGATTAATTAATTAATGATGATGATGATGATGATGATGATG 681
Qy      257 LeuGlnGluGlu---PheAspArgTyIlySerGlyTyIlyTrrPheProIlyAspGluThr 275
Db      682 TATGAAAGAGAAAGTCTTCACTGCTCTGCTGTGTGTGTGTCTTCAAAACGCGACTTTT 741
Qy      276 ThrProSerGlyGlyIlyIleLeuArgIleLeuTyIlyGluGluAsnAspGluSerGluVal 295
Db      742 -----TTAGCATATGCCCAATTTTAACGACACAGAAAGTC 774
Qy      296 GluIleIleHis-----ValThrSerProMetLeuGluThrArgArgAlaAspSer 312
Db      775 CCACCTATTGAATACTCTTCTACTCTGATGATGATGATGATGATGATGATGATGATG 834
Qy      313 PheArgTyIlyProIlyThrGlyThrAlaAsnProIlyValThrPheIlyMetSerGluIle 332
Db      835 GTTCCATATCCAAAGCAGAGCTGTGATTCATTCATGTAAGTTCTTTGTT-----GTA 888
Qy      333 MetIleAspAlaGluGlyArgIleIleAspValIleAspIlyGluLeuIleGlnProPhe 352
Db      889 AATACAGACTCTCTGAGCTGATCCCAATGCAATTCATCAATCAATCAATGCTCTGCT 948
Qy      353 GluIleLeuPheGluGlyValGluTyIlyIleAlaArgAlaGlyTrrPheProGluGlyIly 372
Db      949 TCTATGTTGATA---GGGATCAGACTCTTGTGTATGATGATGATGATGATGATGATG 993
Qy      373 TyralaTrrSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValIleuIleSer 392
Db      994 -----ACACAGAAAGAAATTTCTTTCAGTGGCTC--- 1023

```

QY 393 ProGluLeuPheHeIleProValGluAspAspValMetGluArgGlnIleGluSer 412  
 Db 1024 -----AGAGAGATTGAGAAC 1038  
 QY 413 ValProAspSerValThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsn 432  
 Db 1039 TAT-----TCGGTCATGATATTGTGACCTATGATGATCAATCCAGTGAAGATG--AAC 1089  
 QY 433 IleHisAspIlePheHisValPheProGlnSerHisGluGluGluIleGluPheIlePhe 452  
 Db 1090 TGC-----TTAGTGGCAGCGCAACACATTGAAATGAGTACTACTGCTGCTG 1134  
 QY 453 AlaSerGluCysLysThrGlyPheArgHis-----LeuTyrLys 465  
 Db 1135 GTTGGAGAGATTGAGCTTACAGAACCTCATTTTACCTTGATGTAATACCTTCTACAG 1194  
 QY 466 IleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGlyLysLeuProAlaPro 485  
 Db 1195 ATCATCAGC-----AATGAAGAAAGTTACAGACACATT----- 1227  
 QY 486 SerAspPheLysCysProIleLysGluGluIleAlaIleThrSerGlyLysTyrPheGlu 505  
 Db 1228 TGTATTTCGAAATAGATMAAAAGACTGCACATTTTACAAAGGACCTGGGAAATC 1287  
 QY 506 LeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrPheGlu 525  
 Db 1288 ATCGGG-----ATTGAAGCTCTAACCGATATTATCTACTACTACTATTACT 1332  
 QY 526 GlyThr---LysAspSerProLeuGlnHisIleLeuTyrValValSerTyrValAsnPro 544  
 Db 1333 AATGAATATATAAGAAATGCCAGAGAGAAAGAAATCTTATATAATCCAACTATTGACTAT 1392  
 QY 545 GlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSer---Gln 563  
 Db 1393 ACAAAAGTGCATGACCTC-----AGTTGTGAGCTGAATCCCGAA 1431  
 QY 564 HisCysAspPhePheHeIleSerLysTyrSerAsnGlnLysAsnProHis-----Cys 580  
 Db 1432 AGGTGTACAGTACATTTCTGTGTCATTAGTAAGAGCGCAAGATATTATCGAGTGAATGT 1491  
 QY 581 -----ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 596  
 Db 1492 TCCGCTCTGCTGCTGCCCTCTACTACTACACAGCGCTGATGATTAAGGCTGAGAGA 1551  
 QY 597 ThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 616  
 Db 1552 GTCTCTGGAAGAC--AATTGAGCTTGGATMAAATG-----CTGCAGAAATGTCAGATG 1602  
 QY 617 ProGlu-----IlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 633  
 Db 1603 CCTCCAAAACAACTGAGCTTCATTATTTGATGAATGAACAAATTT---TGTATACAGATG 1659  
 QY 634 LeuTyrLysProHisAspLeuGlnProGlyLysTyrProThrValLeuPheIleTyr 653  
 Db 1660 ATCTGCTCTCTCAT--TTTGATMAATCCAGAAATATCTCTACTATTAGATGTAT 1716  
 QY 654 GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 673  
 Db 1717 GAGAGGCCCATGTAGTCAA-----AAAGCAGACACTGCTTCAAGACTG 1758  
 QY 674 Asn-----ThrLeuAlaSerLeuGlyTyrValValValVal---IleAspAsnArg 689  
 Db 1759 AACTGGGCCACTTACCTTGACAGACAGAAAACATTATAGTACTAGCTTGATGACAGAG 1818  
 QY 690 GlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIle 709  
 Db 1819 GGAAGGTGTTACCAAGAGATTAAGATCATGATCATCAACAGAAAGCTGGGAAACATT 1878  
 QY 710 GluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAsp 729  
 Db 1879 GAAAGTTGAAGATCAATTTGAAGACGACCAACATTT---TCAAAATGGGATTTGTGGAC 1935  
 QY 730 LeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyrIleuSerLeuMetAlaLeu 749

Db 1936 AACAAAGAAATTCATTTGGGCTGCTCATATGAGAGGTAACGTAACCTCAATGCTCTG 1995  
 QY 750 MetGlnArgSerAspIlePheArgValAlaIleAlaGlyValAspValThrLeuTyrIle 769  
 Db 1996 GAGTCGGAGAAATGCGCGTTTCAAGTGTGAGAAATGACCTGGCGCTGTATCCGGTGGAG 2055  
 QY 770 PheTyrAspThrGlyTyrThrGluArgTyrMetGly-----HisProAspGlnAsnGlu 787  
 Db 2056 TACTATGACTCAGTGTACAGAGAACTTACATGGGTCTCCCAACTCCAGAAAGAACCTT 2115  
 QY 788 GlnGlyTyrTyrIleuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsn 807  
 Db 2116 GACCATTCAGAAATTCACAGTCATGACGAGCAGCTGAATTT-----AACAAAGTT 2169  
 QY 808 ArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIle 827  
 Db 2170 GAGTACTCTCTTATTCATGAGAACAGAGATGATATACCTTACCTTACAGCAGTCACTCAG 2229  
 QY 828 LeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGlu 847  
 Db 2230 ATCTCCAAAGCCCTGCTGATGTTGATGGATTTCCAGGCAATGTGTATACTGATGAA 2289  
 QY 848 ArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHisLeuHisTyr 867  
 Db 2290 GACCATGGAATACCTGACGACAGACAGACACCAACATATATATACCAATGAGCCACTTC 2349  
 QY 868 LeuGlnGlu 870  
 Db 2350 ATMAAACAA 2358  
 RESULT 4  
 PCT-US93-07923-1  
 Sequence 1, Application PC/RUS9307923  
 GENERAL INFORMATION:  
 APPLICANT: Morimoto, Chikao  
 APPLICANT: Schloesman, Stuart F.  
 APPLICANT: Tanaka, Yoshiaki  
 TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Fish & Richardson  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
 COMPUTER: IBM PS/2 Model 502 or 55SX  
 OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
 SOFTWARE: WordPerfect (Version 5.0)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/07923  
 FILING DATE: 19930819  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/934,162  
 FILING DATE: 21-AUG-1992  
 APPLICATION NUMBER: 07/832,211  
 FILING DATE: 06-FEB-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fraser, Janis K.  
 REGISTRATION NUMBER: 34,819  
 REFERENCE/DOCKET NUMBER: 00530/005002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 542-5070  
 TELEFAX: (617) 542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2924  
 TYPE: nucleic acid

[illegible]

QY	486	SerPhePheLeuCySPProIleLeuGluIleAlaIleThrSerGlyGluTrpGluVal	505
Db	1163	TGCTATTTTCCAAATAGATMAAAAGACGCACTATTATTACAAAAGGACCGGGAGATC	1222
QY	506	LeuGlyATrGhiLeGlySerAsnIleGluValAspGluValArgGluValTyPheGlu	525
Db	1223	ATCGGG-----ATAGAACTTAAACCAAGTATCTTACTACTACTTACT	126
QY	526	GlyThr---LysAspSerProLeuGluIleHisLeuTyValValSerTyValAsnPro	544
Db	1268	AATGAATATTAAGAAGATGCCAGAGCAAGAAATCTTTATAAATC-----	1311
QY	545	GlyGluValThrArgLeuThrAspArgGlyTySerHisSerCySPeIleSer-----	562
Db	1313	-----CAACTTATGAC-----TATACAAAAGTACATGCTCCATGTTGTAG	135
QY	563	-----GlnHisCySPAspPhePheIleSerLysTySerAsnGluLysAsnProHis	579
Db	1355	CTGAATCCGGAAAGGTGTCAGTACTACTTCTGTGTATTCAGTAAAGAGCGCAAGTATTT	1411
QY	580	-----CyS-----ValSerLeuTyLysLeuSerSerProGluAspAsp	592
Db	1415	CAGCTGAGATGTTCCGCTCTGCTGTCTGCTCCCTCTACTACTACACAGACGCGTGAATAT	1474
QY	593	ProthrCySPLeuThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro	612
Db	1475	AAAGGCGTAGAAGCTCTGAAAGAC--AATTCAGCTTTTGATTAATAATG-----CTGAG	1522
QY	613	AspTyThrProProGlu-----IlePheSerPheGluSerThrThyGlyPheThr	629
Db	1528	AATGTCAGATAGCCCTCCAAAACGTGACCTTCAATTATTGAAAGAAACCAAAATTT---	1588
QY	630	LeuTyGlyMetLeuTyLysProHisAspLeuGlnProGlyLysLysTyProThrVal	649
Db	1583	TGCTATCAGATGATCTTCCTCTCAT---TTTGATTAATATCCAAAGAAATATCTCTACTA	1633
QY	650	LeuPheIleTyGlyGlyProGluValGlnLeuValAsnAsnArgPheLysGlyValLys	669
Db	1640	TTAGATGTGTATGACAGGCCCATGTAGTAA-----AAAGCAGACACT	1681
QY	670	TyrPheArgLeuAsn-----ThrLeuAlaSerLeuGlyTyValValValVal---	685
Db	1682	GTCTTCACAGCTGAACTGGGCGCACTTACTTGCACACAGAAACATTTACTACTAC	1744
QY	686	IleAspAsnArgGlySerCySPHisArgGlyLeuLysPheGluGlyAlaPheLysTyLys	705
Db	1742	TTTGTATGCGAAGAGAGTGTATACCAAGAGATAAGATCATGATGACATCAACAGAGA	1801
QY	706	MetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTy	725
Db	1802	CTGGGACATTTGAAGTTGAAGATCAATTAAGAACAGCCAGCAATTT--TCAAAAATG	1856
QY	726	AspPheIleAspLeuAspArgValGlyLysHisGlyTPSerTyGlyGlyTyLysLeuSer	745
Db	1859	GGATTTTGACACAACAACGAATTCATTTGGGGCTGTCTATATGACGGTACTTAAC	1911
QY	746	LeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProVal	765
Db	1919	TCAATGTGCTCCGGATCAGAAAGTGGCGCTGTTCAAGTGTGAATGTGCCGTGGCGCTGTA	1978
QY	766	ThrLeuTrpIlePheTyAspThrGlyTyThrGlnArgTyPheGly-----HisPro	783
Db	1979	TCCGGTGGAGTACTATGACTCATGCTGTACACAGACCTTACATAGGGCTCCCAACTCA	2038
QY	784	AspGlnAsnGluGlnGlyTyTyLeuGlySerValAlaMetGlnAlaGluLysPhePro	803
Db	2039	GAAAGCAACCTTGACCATTTACAGAAATTCACACAGTCATGAGACAGCTGAAAAATTT---	2095
QY	804	SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla	823
Db	2096	---AAACAAGTTGAGTACTCTCTTATCATGGAACAGACAGATGATTAACGTTTCACTTTGAG	2152

QY 824 HistHsrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrPheLeuGlnIle 843  
Db 2153 CAGTAGCTGCAGATCTCCAAAGCCCTCGATGTTGGATGATTTCCAGCAATGTCG 2212  
QY 844 TyrProGlnIuArgHisSerIleArgValProGluSerGlyLysHisTyrGluLeuHis 863  
Db 2213 TATACTGATGAAGACCATGATAGTACGACACACGACACCAATATATATACCCAC 2272  
QY 864 LeuLeuHisTyrLeuGlnIu 870  
Db 2273 ATGACGCACTTCATAAAACAA 2293  
RESULT 5  
US-08-230-491A-1  
Sequence 1, Application US/08230491A  
Patent No. 5587299  
GENERAL INFORMATION:  
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;  
APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.;  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR  
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FELPE & LYNCH  
STREET: 805 THIRD AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT - ASC II  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/230,491A  
FILING DATE: 20-APRIL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5587299man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 330  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2815 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-230-491A-1  
Alignment Scores:  
Pred. No.: 1,15e-41 Length: 2815  
Score: 454.00 Matches: 177  
Percent Similarity: 41.92% Conservative: 111  
Best Local Similarity: 25.76% Mismatches: 278  
Query Match: 9.66% Indels: 122  
Gaps: 27  
US-10-070-464-1 (1-882) x US-08-230-491A-1 (1-2815)  
QY 212 IleAlaPheIleHisSerAsnAspIleTyrPheSerAsnIleValThrArgGluGluArg 231  
Db 692 TTAGATATAGTCTCAAAACAATATCTTTGAAACAAAGACGAGATCCACCTTTT 751  
QY 232 ArgLeuThrTyrValHisAsnGluLeuAlaAsnMetGluIuAspAlaArgSerAlaGly 251  
Db 752 CAATAAACAATT-----AATGAAAGGAAATAATTAATTTATATGGA 793  
QY 252 ValAlaThrPheValLeuGlnGluIuPhe-----AspArgTyrSerGlyTyrTrpTrp 269

Db 794 ATCCAGACTGGAGTTTATGAAAGAAATGCTTCTACAAATATGCT-----CTCTGCTGG 850  
QY 270 CysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGluGln 289  
Db 851 TCTCCT-----AATGAAATTTTGT-----GCATATGCGGAA 883  
QY 290 AsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGlnThr---Arg 308  
Db 884 TTTAATGATTAAGATATACAGTATGCTTATGCTTATTAATGAGCGATTAACAAATATCT 943  
QY 309 ArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLys 328  
Db 944 AGAACAAATAAATATTCATACCCAAAGCTGAGCTAAGAAATCCCTTTTGGGATA--- 1000  
QY 329 MetSerGluIleMetIleAspAlaGluIuArgIleIleAspValIleAspLysGluLeu 348  
Db 1001 -----TTATATATGATACCACTTACCTGCGTAT 1030  
QY 349 IleGlnProPheGluIleLeuPheGluGluValGluTyrIleAlaArgAlaGlyTrpThr 368  
Db 1031 GTAGGTCCCAGAAAGTGGCTGTTCACGCAATG-----ATAGCCTCAAGTGAATTAT--- 1081  
QY 369 ProGluGlyLysTyrAlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIle 388  
Db 1082 -----TATTCAGTTGGCTCACGGCTTACTGATGAACGATAGTTGGACGTGG 1132  
QY 389 ValLeuIleSerProGluLeuPheIleProValGluAspAspValMetGluArgGlnArg 408  
Db 1133 CTAAAAAG-AGTCCAGAAATGTTCCGCTCCTGTC-----TATATGAGACTTCAGGGA 1182  
QY 409 LeuIleGluSerValProAspSerValThrProLeuIleIleTyr-GluGlnThrThrAs 428  
Db 1183 AGACTGGACAGACATGATGTCCAAGACCA-----GGACATATAGAAAGAAAGCAAGAC 1239  
QY 428 PileTrpIleAsnIleHisAspIlePheHisValPheProGlnSerHisGluGluGluIu 448  
Db 1240 TGGATGGCGCTGT-----GATTTCTTTGTTCAAGACCAAGTTTCAGTATGATGCCAT 1293  
QY 448 eGluPheIlePheAlaSerGluCysLysThrGlyPheArgHisLeuTyrLysIleThrSe 468  
Db 1294 TTCGTACTACAAATATTAGTACAGAGATGATCAAAACATATTCACATAATATC----- 1348  
QY 466 rIleLeuLysGluSerLysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPh 488  
Db 1348 ----- 1348  
QY 488 eLysCysProIleLysGluGluIleAlaIleThrSerGlyGluTrpGluValLeuGluYar 508  
Db 1349 -AAAGACACTGGAAATGCTATTCAAATTCAAGTGGCAAGTGGAGGCCATA----- 1402  
QY 508 HisGluSerAsnIleGlnValAspGluValArgArgLeuValTyrPheGluGlyThrLys 528  
Db 1403 -----AATATA-----TTCAAGATAACAA 1422  
QY 528 AspSerProLeuGlnHisIleLeuTyrValValSerTyrValAsnProGlyGluValTh 548  
Db 1423 GGAATCACTGTTTATTTCTAGCAATGAATTTGAAGAAATCCCTGGAAGAAACATCTTA 1482  
QY 548 rArgLeuThrAspArgGlyTyrSerHisSer---CysCysIleSerGlnHis----- 564  
Db 1483 CAGATTTAGCATGGAAGCTATCTCCAAAGAAAGAGTGTACTGTGCATCTAAGGAA 1542  
QY 565 -----CysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSe 582  
Db 1543 AGAAAGGCGCCAAATTTACACAGCAAGTTTACGCACTACGCAAGTACTATGACCTTGT 1602  
QY 582 rLeuTyr-----LysLeuSerSerProGluAspAspProThrCysValThrLys 598  
Db 1603 CTGCTACGCGCCAGGATCCCATTTCCACCTTATATATGAGCAGCATGATCAAGAA--- 1660  
QY 598 sGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp----- 613

```

Db      1661 -----ATTAATAATCTGGAGAGAAAAAGAGATTGGAAATGCTTTGAAAAATAT 1710
Qy      614 -TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrIleuTyrclyme 633
Db      1711 CCAGCTGCTTAAGAGAAATTAAGAACTTGAAGTAGATGAATTAATTAAGTAGACA 1770
Qy      633 tLeuTyrclyProHleAspLeuGlnProGlyIleTyrclyProThrValLeuPheIleTy 653
Db      1771 GATGATTCCTCTCCCAATTGACAGATCAAGAAAGATACCTTCCTTAATTAAGTGA 1830
Qy      653 rGlyGlyProGlnValGlnLeuValAsnAsnArgPheIleValIleTyrclyPheArgle 673
Db      1831 TGGTGCTCCCTGACAGTCAAGTGTAAAGTCT-----GTATTTCCTGT 1872
Qy      673 uAsnThrIleuAlaSerLeu-----GlyTyrclyValValValIleAspAsnArg 699
Db      1873 TAAATGGATATCTTATCTTTCGAAGTAGAGAGGATGGTCAATTCCTTCCTGGTGGATGTTG 1932
Qy      689 gGlySerCysHleAspGlyLeuIleAspPheGluGlyAlaPheIleTyrclyIleMetGlyGln 1709
Db      1933 AGGAACAGCTTCCAAAGTGACAAACTCTCTATGCAAGTATCGAAAGCTGGGTGTTTA 1992
Qy      709 eGluIleAspArgIleValGlnGlyLeuGlnIleuAlaSerArgTyrclyAspPheIleAs 729
Db      1993 TGAAGTTGAAGACCAAGATTACAGCTGTCAAGAAATTCATA--GAATGGGTTTCATTGA 2049
Qy      729 pLeuAspArgValGlyIleHleGlyTyrclyProGlyIleTyrclyLeuSerLeuMetAlaIle 749
Db      2050 TGAATAAAGATAGCCATATGGGGCTGTCTATGAGAGATACGTTTCACTCACTGGCCCT 2109
Qy      749 uMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrIleuThrIle 769
Db      2110 TCCATCTGGAACGTGCTTTTCAAAATGTGTATAGCAAGGCTCCAGTCTCCAGCTGGGA 2169
Qy      769 ePheTyrclyAspThrGlyTyrclyThrGluArgTyrclyMetGlyHlePro-----AspGlnAsnG 787
Db      2170 ATATTACCGGCTGTCTACACAGAGAGATTCAAGGCTCTCCCAAGAGATGATTAATCT 2229
Qy      787 uGlnGlyTyrclyLeuGlySerValAlaMetGlnAlaGlyIlePheProSerGluProAs 807
Db      2230 TGAGCACTTAAGAAATTCAGCTGTGTGCGCAAGAGCAATATTTTCAGAAAGTAGTAC-- 2287
Qy      807 nArgLeuLeuLeuLeuHleGlyPheLeuAspGlnAsnValHleAspAlaHleThrSerIle 827
Db      2288 ----TATCTTTCATCCACAGCAAGACATATATGTGCACCTTCAAAACTCAGACACA 2343
Qy      827 eLeuLeuSerPheLeuValArgAlaGlyIleProTyrclyAspLeuGlnIleTyrclyProGln 847
Db      2344 GATTCCTAAAGCTGTGTTAATGACAAAGTGAATTCAGGCAATGTGTACTCTGACCA 2403
Qy      847 uArgHleSerIleArgValProGluSerGly-----GluHleTyrclyLeuHleIle 864
Db      2404 GAACCAACGGCTTA-----TCCGGCTGTCCAGAAACCACTTATACACCCACAT 2451
Qy      864 uLeuHleTyrclyLeuGlnIle 870
Db      2452 GACCCACTTCTTAAGCAG 2470

```

RESULT 6  
US-08-619-280A-1  
Sequence 1, Application US/08619280A  
Patent No. 5767242

GENERAL INFORMATION:  
APPLICANT: Zimmermann, Rainer; Park, John E.;  
Applicant: Rettig, Wolfgang; Old, Lloyd J.  
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN  
TITLE OF INVENTION: ALPHA, AND USES THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York

```

/ COUNTRY: USA
/ ZIP: 10022
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
/ COMPUTER: IBM PS/2
/ OPERATING SYSTEM: PC-DOS
/ SOFTWARE: Wordperfect
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/619,280A
/ FILING DATE: 18-MARCH-1996
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/230,491
/ FILING DATE: 20-APRIL-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hanson, No. 5767242man D.
/ REGISTRATION NUMBER: 30,946
/ REFERENCE/DOCKET NUMBER: LUD 5330.1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 688-9200
/ TELEFAX: (212) 838-3884
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2815 Base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/
/ US-08-619-280A-1
/
/ Alignment Scores:
/ Pred. No.: 1,15e-41 Length: 2815
/ Score: 454.00 Matches: 177
/ Percent Similarity: 41.928 Conservative: 111
/ Best Local Similarity: 25.768 Mismatches: 278
/ Query Match: 9,668 Indels: 122
/ Gaps: 27
/
/ US-10-070-464-1 (1-882) x US-08-619-280A-1 (1-2815)
/
Qy      212 IleAlaPheIleHisSerAsnAspIleTrpIleSerAsnIleValThrArgGluIleArg 221
Db      692 TTAGCATATGTCATCAAAACAATATCTATTGAAACAAAGACAGAGATCCACCTTTT 751
Qy      232 ArgLeuThrTyrclyValHleAsnGlnLeuAlaAsnMetGluGlnAspAlaArgSerAlaGly 251
Db      752 CAATTAACATTT-----AATGGAAGAGAAATTAATTAATGGA 793
Qy      252 ValAlaThrPheValLeuGlnGluPhe-----AspArgTyrclyGlyTyrclyTrp 269
Db      794 ATCCCAAGCTGGGTATATGAGAGAGGAATGCTTCTCAAAATATGCT--CTGTGATG 850
Qy      270 CysProIleAlaGluThrThrProSerGlyGlyIleLeuValIleLeuArgIleLeuTyrclyGlu 289
Db      851 TCTCCT-----AATGGAATTTTGG-----GCATATGGGAA 883
Qy      290 AsnAspGluSerGluValGluIleHleIleValHleThrSerProMetLeuGlnThr--Arg 308
Db      884 TTTAATGATTAAGATATACCACTTATGCTTATTCCTATTATGCGCATGAACAATATCTCT 943
Qy      309 ArgAlaAspSerPheArgTyrclyProIleThrGlyIleThrAlaAsnProIleValHlePheIle 328
Db      944 ACMAACAATTAATATTCATACCAAGGCTGAGCTTAAGAAATCCGCTGTGCGATA--- 1000
Qy      329 MetSerGluIleMetIleAspAlaGluIleArgIleIleAspValIleAspIleGluIleu 348
Db      1001 -----TTATATGATGATACCACTTACCTGCTAT 1030
Qy      349 IleGlnProPheGluIleLeuPheGluGlyValGluTyrclyAlaArgAlaGlyTyrclyThr 368
Db      1031 GTAGGTCCCAAGAGTCTGTTCACCAATG-----ATAGCCTCAAGTATTAAT--- 1081
Qy      369 ProGluGlyTyrclyAlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIle 388

```



```

Db      1082 -----TATTTAGTTGGCTCAGTGGGTTACTGATGACGAGNATGTTGCAGTGG 1132
Qy      389 ValLeuILSerProGluLeuPheIleProValGluAspAspValMetGluValGlnIleA 408
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1133 TAAAAAAG-AGTCCAAATGTTTCGGCTCGTC-----TATATGCACTTCAGGGA 1182
Qy      409 LeuILGluSerValProAspSerValThrProLeuIleIleTyr-GluGluThrIleAs 428
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1183 AGACTGCACACATGGGATTTGCCAAAGACCA--GGACATATAGAAACAAAGCAGAAC 1239
Qy      428 pIleTPriLeAsnIleHisAspIlePheHisValPheProGlnSerHisGluGluGlu 448
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1240 TGGATGGGCTGCT-----GGATTCCTTTGTTTCAAGCCAGTTTCAGCTATGATGCAT 1293
Qy      448 eGluPheIlePheAlaSerGluCysIleThrGlyPheArgHisLeuTyrIleThrSe 468
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1294 TTCGTACTCAAAATTTTATAGTACAAAGATGGCTCAACAAATTCACATATTC----- 1348
Qy      468 rIleLeuLysGluSerLysTyrLysArgSerSerGlyLeuProIleProSerAsp 488
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1348 ----- 1348
Qy      488 eLysCysProIleLysGluGluIleAlaIleThrSerGlyLysIleProGluValGluGly 508
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1349 -AAAGACACTGTGCAAAATGCTATTCAAATTACAGTGCAGAGTGGAGGCCAT- 1402
Qy      508 gHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrPheGluGluThrIly 528
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1403 -----AATATA-----TTCAAGACTATACACA 1422
Qy      528 sAspSerProLeuGluHisIleHisLeuTyrValIleSerTyrValAsnProGlyGluValTh 548
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1423 GGATTCACGTTTATTTCTAGCATGATTTGAAGAATTCCTCGAAGAAACATCTTA 1482
Qy      548 rArgLeuThrAspArgGlyTyrSerHisSer---CysCysIleSerGlnHis----- 564
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1483 CAGAAATTAGCATTTGAAAGCTATCTCCACAGCAAGATGGTACTTCGATCTTAAGAA 1542
Qy      565 -----CysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSe 582
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1543 AGAAAGTGCCTAATATTACACACAGCAAGTTTCAGAGCACTCCCAAGTACTATCACCTGT 1602
Qy      582 rLeuTyr-----LysLeuSerSerProGluAspAspProThrCysIleThrIly 598
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1603 CTGCTACGGCCGACGAGTCCCATTTCCACCTTCATGATGAGCAGCATGATCAGAA-- 1660
Qy      598 sGluPheTPriAlaThrIleLeuAspSerAlaGlyProLeuProAsp----- 613
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1661 -----ATTAAATCTCGAGAGAAACAAAGAAATGGAAATGCTTTGAAAAATAT 1710
Qy      614 -TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrIleuTyrGlyMe 633
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1711 CCAAGCTGCCTTAAGAGGAATTTAAGAACTTGAAGTGAAGAAATTACTTTATGATGTA 1770
Qy      633 rLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyr 653
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1771 GATGATTTCTTCTCTCAATTTGACAGATCAAAAGATATCCCTTGCTATCAAGTGA 1830
Qy      653 rGlyIleProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgIle 673
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1831 TGGTGGTCCCTGCAGTGCAGGTGAAGTCT-----GTAATTTGCTGT 1872
Qy      673 uAsnThrLeuAlaSerLeu-----GlyTyrValValValValIleAspAsnArg 689
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1873 TAATTGGATATCTTATCTTGCAGTAAAGAAAGGATGCTATTCCTTGCTGATGATGCTG 1932
Qy      689 gGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnI 709
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1933 AGGAAAGACTTTTCAAGGTGACAAACTCTCTATGACAGTATGCAAAAGCTGGGTGTTTA 1992
Qy      709 eGluIleAspAspGluValGluGluLeuGlnTyrLeuAlaSerArgTyrAspPheIleAs 729
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1993 TGAAGTTGAAGACCAATTTACAGCTGTCAAGAAATTCAT- 2049

```

```

Qy      729 rLeuAspArgValGlyIleHisGlyThrSerTyrGlyGlyTyrLeuSerLeuMetAlaIle 749
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2050 TGAATAAAAGATATGCCATATGGGCTGTCTATGAGAGATATGCTTATCATCTGGCCCT 2109
Qy      749 uMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTPriI 769
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2110 TGCATCTGCAACCTGCTTTTCAATATGCTGTATAGCAGAGTGGCTCCAGCTCCAGCTGGGA 2169
Qy      769 ePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro-----AspGlnAsnG 787
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2170 ATATTACGGCTGTCTGTACACAGAGATTCATGAGTGTCTCCCAACAAAGATGATAATCT 2229
Qy      787 uGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAs 807
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2230 TGAGCACTATTAAGAAATTCAACTGTGTGTCAGAAAGCAGAAATATTTAGAAATGTAAC- 2287
Qy      807 rArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerI 827
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2288 -----TATCTTCTCATCTCACAGAACACAGACATGATTAATGCTTCAAAACACAGACA 2343
Qy      827 eLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnG 847
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2344 GATTGCTAAAGCTCTGTTAATGCACAAGTGAATTCAGAGCAATGTGTACTGTGACCA 2403
Qy      847 uArgHisSerIleArgValProGluSerGly-----GluHisTyrGluLeuHisIle 864
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2404 GAACCAAGGCTTA-----TCCGGCTGTCCACAGAACCACTTATACACCCACAT 2451
Qy      864 uLeuHisTyrLeuGlnGlu 870
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2452 GACCACTTCTTAAGCAG 2470

```

RESULT 7  
 US-08-940-391-1  
 ; Sequence 1, Application US/08940391  
 ; Patent No. 5965373  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zimmermann, Rainer; Park, John E.;  
 ; TITLE OF INVENTION: Retig, Wolfgang; Old, Lloyd J.  
 ; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; STREET: 805 Third Avenue  
 ; CITY: New York City  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10022  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: PC-DOS  
 ; SOFTWARE: Wordperfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/940,391  
 ; FILING DATE: 01-OCT-1997  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/619,280  
 ; FILING DATE: 18-MARCH-1996  
 ; APPLICATION NUMBER: 08/230,491  
 ; FILING DATE: 20-APRIL-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hanson, No. 5965373man D.  
 ; REGISTRATION NUMBER: 30,946  
 ; REFERENCE/DOCKET NUMBER: LUD 5330.1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 688-9200  
 ; TELEFAX: (212) 688-3884  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-940-391-1

Alignment Scores:
Pred. No.: 1.15e-41 Length: 2815
Score: 454.00 Matches: 177
Percent Similarity: 41.92% Conservative: 111
Best Local Similarity: 25.76% Mismatches: 278
Query Match: 9.66% Indels: 122
DB: 2 Gaps: 27

US-10-070-464-1 (1-882) x US-08-940-391-1 (1-2815)

QY 212 ILEAIAPIHEIHSERLNSAPRIETRIPILESERAMILEVALIHRGGLUGIATG 231
DB 692 TTAGCATATGTCCTACCAACAATATCTATTTGAAACCAAGACAGAGATCCACCTTTT 751
QY 232 ARGLEUTHIRYVALHISANGIULEUALAASMETGLUGIUAAPALARGSERIAGIY 251
DB 752 CAAARACATTT-----AATGAGAGAAATAATAATTATTATGGA 793
QY 252 VALAIAHRPHEVALLEUNGINGIUPHE-----ASPARGTYSERGIYTRITRP 269
DB 794 ATCCGACAGCTGGTTATGAGAGAAATGCTTCTACAAATAATGCT---CTCGGTGG 850
QY 270 CYSPIROLYSIAGLIUTHIRPROSERGIYLYVILEUARGILEUTHYRGLUGIU 289
DB 851 TCTCTCT-----AATGAAAAATTTTGT-----GCATATGCGGAA 883
QY 290 ASPAERGUSERGLUVALIGIUILEIHSIVALIHRSERPROMETLEUNGUTHR---ARG 308
DB 884 TTTATGATTAAGATATACCACTTATTCCTATTATGGGATGAAACAATATCTT 943
QY 309 ARGALAAERSEPRHEARGTYRPIROLYSIRGIYTRIALAASPIROLYEVALIHRPHELYS 328
DB 944 AGAACATAAATATATCCATACCCAAAGCTGAGTAAGAAATCCCGTGTTCGATA--- 1000
QY 329 METERGULIEMERIIASAALAGIUGIYARGIILEASPRVALIIEASPLYSGIULEU 348
DB 1001 -----TTTATTTATCGATATCCACCTTACCTCGGTAT 1030
QY 349 ILEGINPROPHEGULILEUPHEGLUGIYVALIGIUYTRILEAARGALAGIYTRIPHR 368
DB 1031 CTAGAGTCCCAAGAAAGTCCCTGCCACAAATG-----ATAGCTTCAAGATATAT--- 1081
QY 369 PROGLUGIYSTRYRATATPISERILEUUEHAPARGSERGINTHIRARGLEUNGILE 388
DB 1082 -----TATTTCAAGTTGGCTCAGTGGGTACTGATGAACGAGTATGTTGCAATGG 1132
QY 389 VALLEUIESERPROGLIULEUPHELEPROVALIGIUAERPAVALMETGLUGIARG 408
DB 1133 CTAAAAAG-ACTCCAGAAATGTTTGGTCTGTC-----TATATGTGACTTCAGGGA 1182
QY 409 LEUILEGUSERVALIPROASPRSERVALIHRPROLEUILEIETRYR-GIUGIUTHRTHAS 428
DB 1183 AGAGTGGCAGACATGGGATTTCCAAAGACCCA---GGAGCATATAGAAAGAACAGAAC 1239
QY 428 PIETRIPILESANILEHISAPRIEHEHISVALIHRPHEGINSEHISGLUGIUI 448
DB 1240 TGGATGGGCTGGT-----GGATTCCTTTGTTTCAAGACAGATTTTCAGCTAGTGCAT 1293
QY 448 EGLUPHEIIEHIEHISERGIUCYSLYRTHGIYRPHARGHISLEUTHYRYSIIEHRSE 468
DB 1294 TTGCTACTACAAATAATTTAGTGACAAAGATGGCTACAAACATATTTCACTATATC--- 1348
QY 468 RILEULEYGLUSERIYSTRYLSARGSERSEGIYGIYLEUPROALAPROSEARPH 488
DB 1348 ----- 1348
QY 488 ELYSCSPROIIEYSGIUGIUILEAIAIETHRSEGIYGIUTRPGIUAILEUGIYAR 508

```

```

DB 1349 -AAAGCACTGGAGAAATGCTATTCAATTTACAAATGGCAATGGAGGCCATA----- 1402
QY 508 GHISGLYSERAPRIIEGLIIVALSPGIUVALIARGARGLEUVALIYRPHIGIUIYTHRY 528
DB 1403 -----AATATA-----TTCAGAGTAACACA 1422
QY 528 SAPSERPROLEUGIUIHEIHSERLEUTHYRVALIASERTYRVALAASPROGLIUIVALTH 548
DB 1423 GGAATTCAGCTGTTTATTTATTCAGAAATGAATTTGAAGATCCCTGGAAAGAAACCACTTA 1482
QY 548 RARGLEUTHIRAPARGIYTYRSEHISER---CYSCYIIESEGINHIS----- 564
DB 1483 CAGAAATTAGCATTTGGAAGCTATCCCTCCAAAGAAAGATGTGTACTTGCATTTAAGAA 1542
QY 566 -----r-CYASPRPHEPIIESERIYETYSERANGIULYASPROHISCYVALISE 582
DB 1543 AGAAAGTGTCCAAATATTACACAGCAAGTTTCAGCGCACTACGCAAGTACTATGCACTTGT 1602
QY 582 RLEUTHYR-----LYSEUSERSEPRGIIUAERAPPROTHCYELYSTHRY 598
DB 1603 CTGCTACGGCCAGGCATCCCATTTCCACCTTCATGATGAGCGCATGATCAAGAA-- 1660
QY 598 EGLUPHETRALATHRIIELEUASPRSERIAGIYPROLEUPROASP----- 613
DB 1661 -----ATTAAATTCCTGGAAAGAAACAGAAATTTGAAATGCTTTGAAATAAT 1710
QY 614 -TYRTHRPROPROGLIUIERHESEPRHEGLUSERTHIRGIIYRPHETRYRGIYME 633
DB 1711 CCAGCTGCCTAAAGAGAAATTAAGAACTTGAAGTATGAAATTAATTACTTTATGATACAA 1770
QY 633 CLEUTHYRSPROIHSAPREUGINPROGIYLYVILYSTYRPROTHRVALLEUPHEIIEFY 653
DB 1771 GATGATTTCTTCTCTCAATTGACAGATCAAAAGAAATATCCCTTGATTAATCAAGTGTA 1830
QY 653 RGIYGIYPROGINVALIGIULEUVALIASAPARPHELYSGIYVALIYSTYRPHARGIE 673
DB 1831 TGGTGTCTCCCTGACGTCAAGGTATGAGTCT-----GTAATTTGCTGT 1872
QY 673 UASNTHRIEULASERLEU-----GIYTRVALIVALIIEASPARNAR 689
DB 1873 TAATGATATATCTTATCTTGCAAGTAAGAAAGGATGGCATGCTTGCTGGTATGATGCTG 1932
QY 689 GGLYSERCYSHISARGIYLYLEUPHEGLUGIYALAPHELYSTRYLYSMETGICLINI 709
DB 1933 AGAACAAGCTTTCCAAAGGTGACAAATCTCTATAGCACTGATCCGAAAGCTGGGTGTTA 1992
QY 709 EGLUIEASPARGINVALIGIUGIYLEUNGITRYRLEULASERARGTYRASPHEIIEAS 729
DB 1993 TGAAGTTGAACACAGATTAACGCTGTCAAAATTCATA---GAATGGGTTTCAATGA 2049
QY 729 PLEUASPARGVALIGIUILEHISGLIYTRPSETRYRGIYTYRLEUSERLEUMETALIE 749
DB 2050 TGAATAAAGAAATAGCATATGAGGCTGGCTCTATAGAGATATGCTTATACACTGGCCCT 2109
QY 749 UMETGLIARGSERAPRIEPRHEARGVALIALLIIEALAGIYALAPROVALIHRLEUTHRI 769
DB 2110 TGCATCTGGAACGTGCTTTTCAATGTGTATAGCACTGCTCCACAGTCTCCAGCTGGGA 2169
QY 769 EPHEIYASPRTHRGIYTRHGIUARGTYRMEGIYHISPRO-----APRGIIASGI 787
DB 2170 ATATTAGCGCTGTCTACACAGAGATTTATGAGGCTTCCCAAAAGATGATTAATCT 2239
QY 787 UGLINGIYTYRILEUNGISERVALIAMEGINALIGIULYSPHEPROSEGIUPROAS 807
DB 2230 TGAACCATATAGAAATCACTGATGAGCAAGCAAGCAAGAAATTTCCAGAAATGTAGAC-- 2287
QY 807 NARGLEULEULEUWHISGLIYRHELEUASPRGIASNAIHSIAPHEIAHISHRSEI 827
DB 2288 ---TATCTTCTATCCACGAGAACAGCATGATGAATGTGACTTTCAAAACTCAGACA 2343
QY 827 ELEULEUSERPHEUVALIARGIAGIYLYSPROTYRASPLEUNGILEYRPROGINGI 847

```

Db 2344 GATTGCTAAGCTCTGTTAATGACACAGATGATTTCCAGGCAATGTGTAATCTGACCA 2403  
Qy 847 uArHisSer11eArGValProGluSerGly-----GluHisTyrGluLeuHisLe 864  
Db 2404 GAACCCAGCGCTTA-----TCCGCGCTGTCCAGCAACCACTTAATACACCACAT 2451  
Qy 864 uLeuHisTyrLeuGlnGlu 870  
Db 2452 GACCCACTTCTTAAGCAG 2470

RESULT 8  
US-09-392-184-31/c  
Sequence 31, Application US/09392184  
Patent No. 6395889  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN  
FILE REFERENCE: 5800-55  
CURRENT APPLICATION NUMBER: US/09/392,184  
CURRENT FILING DATE: 1999-09-09  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 31  
LENGTH: 612  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(612)  
OTHER INFORMATION: prolyl oligopeptidase  
NAME/KEY: misc\_feature  
LOCATION: (1)..(612)  
OTHER INFORMATION: n = A,T,C or G  
US-09-392-184-31

Alignment Scores:  
Pred. No.: 4,91e-37 Length: 612  
Score: 404.00 Matches: 97  
Percent Similarity: 69.28% Conservative: 9  
Best Local Similarity: 63.40% Mismatches: 45  
Query Match: 8.60% Indels: 4  
Gaps: 0  
DB: 4

US-10-070-464-1 (1-882) x US-09-392-184-31 (1-612)

Qy 710 Glu11eAspAspGlnValGluGlyLeu-GlnTyrLeuAlaSerArgTyrAspPhe11eAs 729  
Db 519 GAATTGTCGATCAGCGTGGGNGGACTCCCAATATCTAACTTCGAAATGATTCATGTA 460  
Qy 729 Pleu-AspArgValGly11eHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaL 749  
Db 459 CTTAAATGATGATGCGAATCCNGCGCTG-TCCATAGAGGATACCTCTCCCTGATGGCAN 401  
Qy 749 eumecGlnArgSerAsp11ePheArgValAla11eAlaGlyAlaProValThrLeuTrr1 769  
Db 400 TAATTAGAGCTCAGATATCNCAGGGTCT-ATTCTGGGGCCCAAGTCACCTCTGCA 342  
Qy 769 1ePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnG 789  
Db 341 TCTTCTATGAAACAGGAAACNCGGGAGCTTANATGGGTCACTTNNCCAGNATGACAGG 282  
Qy 789 1TyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgL 809  
Db 281 GCTATTACTTAGATCTGGGCTATTCACAGCAAGAAAGTTCCTCTGNAACCAATNNGT 222  
Qy 809 euleuleuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSer11eLeuL 829  
Db 221 TCTTCTCTTANATGGTTCTCTGATTTAGATTTCANNTTTCACANACCAAGNANNTAC 162  
Qy 829 euseSerheLeuValAlaGlyAlaGlyLysProTyrAspLeuGlnLeuTyrProGlnGluArgH 849  
Db 161 TTAGGTTTATGAGGCTGGGAAACCAATGANTTNCAGAACCAACCTCNGAGAGAGN 102

Qy 849 isSer11eArGValProGluSerGlyGluHisTyr 860  
Db 101 CCAACANNAGAGTCTGTAATCGGAGGACAAATAT 67

RESULT 9  
US-08-699-103B-1/c  
Sequence 1, Application US/08699103B  
Patent No. 6107462  
GENERAL INFORMATION:  
APPLICANT: Rine, Jasper D.  
TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 2200 Sand Hill Road, Suite 100  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: IBM Compatible  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: 16-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/002,581  
FILING DATE: 17-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D.  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 09272/005001  
TELEPHONE: 650/322-5070  
TELEFAX: 650/854-0875  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4982 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-699-103B-1

Alignment Scores:  
Pred. No.: 5,29e-28 Length: 4982  
Score: 339.50 Matches: 81  
Percent Similarity: 52.07% Conservative: 32  
Best Local Similarity: 37.33% Mismatches: 81  
Query Match: 7.22% Indels: 23  
Gaps: 7  
DB: 3

US-10-070-464-1 (1-882) x US-08-699-103B-1 (1-4982)

Qy 646 TyrProThrValLeuPhe11eTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPhe 665  
Db 4941 TATCTCGTATTTTCTTTCGATATGCGGACCGAATTCACCAAGTT----- 4894  
Qy 666 LysGlyValLysTyrPheArgLeuAsn-----ThrLeuAlaSerLeuGlyTyr 681  
Db 4893 -----GTCAAAACGTTTCCGTAGATTATTAAGAGTGAGTTCACCAATTAACGCA 4840  
Qy 682 ValValValVal11eAspAsnArgGlySerGlyHisArgGlyLeuLysPheGluGlyAla 701  
Db 4839 ATGTAGTGTGTGTGACGCGTGTGTACTGCTTCAAGAGTTCACCACTTTAGATCCCTT 4780  
Qy 702 PheLysTyrLysMetGlyGlnLeuGluLeuAspAspGlnValGluGlyLeuGlnTyrLeu 721  
Db 721

```

Db      4779 GTTCGGATAGGCTGCTGATACGAGCCCGCAGCAAAATATCTGCGGCTTCTTATAT 4720
Qy      722 AAlaserArGTYrAspRheleuAspArGValAGlylleHleAGlyTTrpSerTyrgly 741
Db      4719 GGTTCCT---TTAACTTTTGTGATCCGCAAAAGATTTCTTATTTGGTGGTCATACGGG 4663
Qy      742 GlyTyLeuSerLeuMetAlaleuMetGlnArGSerAsp---llePheArGValAlaile 760
Db      4662 GCGTACCTGACACTAAAACTTTGGAGAAAGATGGCGGAGACATTTCAAATACCGGATG 4603
Qy      761 AlaGlyAlaProValThrLeuTrpIlePheTyAspRhglyTyTrhGlnArGTYrMet 780
Db      4602 TCAGTTGGCCGACGTAAACGACGAGATTTCAGATTCTGTTTAACTGAGAGGTACATG 4543
Qy      781 GlyHisProAspGlnAsnGlnGlnGlyTyTyTyLeuGlySerVal-----Ala 796
Db      4542 CATACTCCTCAAGAAACTTTCATGATGATACGATACAGGCTTCATATGCTACTGCT 4483
Qy      797 MetGlnAlaGlyLysPheProSerGluProAsnArGLeuLeuLeuHleAGlyPheLeu 816
Db      4482 TTG-----GCACAGCAAAATAGATTGTTTGTGATGACGAGCAAGGA 4441
Qy      817 AspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArGAlaGly 836
Db      4440 GATGATACGTTCACTTCAAAATTCCTTAAAGTTTCGACCTTTTGATCTAAATGCT 4381
Qy      837 ---LysProTyAspLeuGlnIleTyProGlnGlnIuArGHisSerIleArg 852
Db      4380 GTGAAAAATTAATGACGTCCACGCTTCTCTGACTGACATCATATATAGA 4330

RESULT 10
US-09-229-059-1/c
; Sequence 1, Application US/09229059
; Patent No. 6333172
; GENERAL INFORMATION:
; APPLICANT: Rine, Jasper D.
; APPLICANT: Hampton, Randolph
; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/229,059
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,103
; FILING DATE: 16-AUG-1996
; APPLICATION NUMBER: 60/002,581
; FILING DATE: 17-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 09272/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/322-5070
; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4982 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: Genomic DNA
US-09-229-059-1
Alignment Scores:
Pred. No.: 5,296-28
Score: 339,50
Percent Similarity: 52.07%
Best Local Similarity: 37.33%
Query Match: 7,22%
DB: 4
Gaps: 7

US-10-070-464-1 (1-882) x US-09-229-059-1 (1-4982)
Qy      646 TyrProThrValIleuPheIleTyrglyGlyProGlnValAGlnLeuValAsnAsnRgPhe 665
Db      4941 TATCTGATATTTTCTTTCATATATGGGAGCGAATTCATCAAGTT----- 4894
Qy      666 LysGlyValLysTyPheArGLeuAsn-----ThrLeuAlaSerLeuGlyTyR 681
Db      4893 -----GTCAAACGTTTCCGTAGGATTTATGAAAGGTAGCTTCAATTAACCA 4840
Qy      682 ValValValValIleAspAsnArGlySerCyshIsArGlyLeuLysPheGlyVala 701
Db      4839 ATTGATGTTGTTGTTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4780
Qy      702 PheLysTyTyLysMetGlyGlnIleGlylleAspAsnGlnAGlyLeuGlnTyTyLeu 721
Db      4779 GTTCGGATAGGCTGCTGATACGAGCCCGCAGCAAAATATCTGCGGCTTCTTATAT 4720
Qy      722 AAlaserArGTYrAspRheleuAspArGValAGlylleHleAGlyTTrpSerTyrgly 741
Db      4719 GGTTCCT---TTAACTTTTGTGATCCGCAAAAGATTTCTTATTTGGTGGTCATACGGG 4663
Qy      742 GlyTyLeuSerLeuMetAlaleuMetGlnArGSerAsp---llePheArGValAlaile 760
Db      4662 GCGTACCTGACACTAAAACTTTGGAGAAAGATGGCGGAGACATTTCAAATACCGGATG 4603
Qy      761 AlaGlyAlaProValThrLeuTrpIlePheTyAspRhglyTyTrhGlnArGTYrMet 780
Db      4602 TCAGTTGGCCGACGTAAACGACGAGATTTCAGATTCTGTTTAACTGAGAGGTACATG 4543
Qy      781 GlyHisProAspGlnAsnGlnGlnGlyTyTyTyLeuGlySerVal-----Ala 796
Db      4542 CATACTCCTCAAGAAACTTTCATGATGATACGATACAGGCTTCATATGCTACTGCT 4483
Qy      797 MetGlnAlaGlyLysPheProSerGluProAsnArGLeuLeuLeuHleAGlyPheLeu 816
Db      4482 TTG-----GCACAGCAAAATAGATTGTTTGTGATGACGAGCAAGGA 4441
Qy      817 AspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArGAlaGly 836
Db      4440 GATGATACGTTCACTTCAAAATTCCTTAAAGTTTCGACCTTTTGATCTAAATGCT 4381
Qy      837 ---LysProTyAspLeuGlnIleTyProGlnGlnIuArGHisSerIleArg 852
Db      4380 GTGAAAAATTAATGACGTCCACGCTTCTCTGACTGACATCATATATAGA 4330

RESULT 11
US-09-628-133-1/c
; Sequence 1, Application US/09628133
; Patent No. 6531292
; GENERAL INFORMATION:
; APPLICANT: Rine, Jasper D.
; APPLICANT: Hampton, Randolph
; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA

```

```

/ ZIP: 94025
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ OPERATING SYSTEM: IBM Compatible
/ SOFTWARE: FASTSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/628,133
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/699,103
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Green, Grant D.
/ REGISTRATION NUMBER: 31,259
/ REFERENCE/DOCKET NUMBER: 09272/005001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650/322-5070
/ TELEFAX: 650/854-0875
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4982 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA
/ US-09-628-133-1

Alignment Scores:
Pred. No.: 5,29e-28 Length: 4982
Score: 339.50 Matches: 81
Percent Similarity: 52.07% Conservative: 32
Best Local Similarity: 37.33% Mismatches: 81
Query Match: 7.22% Indels: 23
DB: 4 Gaps: 7

US-10-070-464-1 (1-882) x US-09-628-133-1 (1-4982)
QY 646 TyrProThValIleuPheIleTyrgIyGlyProGlnValIleuValAsnAsnArgPhe 665
DB 4941 TATCGTATTTCTTTCATATGCGGACCGAATTCACAAAGTT-----4894
QY 666 LysGIValIleTyPheArgLeuAsn-----ThleuAlaSerLeuGIYr 681
DB 4893 -----GTCMAAACGTTTCCGTAGATTAAAGAGTGAGTCTTCAACATTAACGCA 4840
QY 682 ValValIleValIleAspAsnArgIleSerCyHisArgIleuIlePheGluGlyAla 701
DB 4839 ATTGATGTTGTTGATGCGTGTGCTGCTGCTTCAAGGTCAGACATTAGATCCCTT 4780
QY 702 PheIleTyLeuMetGIyGlnIleGluIleAspAsnGlnValIleGlyLeuGlnTyLeu 721
DB 4779 GTTCGGATAGGCTGGTATTAACGAGCGCCGACCAAAATATCTGCGCTTCTTATAT 4720
QY 722 AlaSerArgTyAspPheIleAspLeuAspArgValGlyIleHisGIYrTyrSerTyrgIy 741
DB 4719 GGTTCCT--TTAACTTTTGTATCCGCAAAAGATTTCCTTATTTGGTGTGTCATACGGG 4663
QY 742 GlyTyLeuSerLeuMetAlaLeuMetGlnArgSerAsp--IlePheArgValAlaIle 760
DB 4662 GGGTACCTGACCTAAAACTTTGAGAAAGATGGCGGAAGCATTTCAAAATACGGGATG 4603
QY 761 AlaGlyAlaProValThIleuTyrIlePheTyAspThrGIYrTyrThGlnArgTyrmec 780
DB 4602 TCAGTTGGCCGAGTAACGAGCTGAGATTTTACGATTTCTGTTATACGAGAGATACG 4543
QY 781 GlyHisProAspGlnAsnGluGlnGlyTyTyLeuGlySerVal-----Ala 796
DB 4542 CATACTCTGCAAGAAACTTTGATGATGATCAAGGCTTCATTAATGTCACTGCT 4483
QY 797 MetGlnAlaGluTyPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 816
DB 4482 TTG-----GCACAGCAAAATATGATTTTGTATGATGACGGAACAGGA 4441

```

```

QY 817 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 836
DB 4440 GATGATACGTTACCTTCAAAATTCCTTAAGCTTTGAGACTTTGATCTAAATGGT 4381
QY 837 ---LysProTyAspLeuGlnIleTyPyrGlnGlnIleArgHisSerIleArg 852
DB 4380 GTGAAATATGACGCTCCACTTCTGACTGACATCATATGATAAGA 4330

RESULT 12
US-09-221-017B-646/C
/ Sequence 646, Application US/09221017B
/ Patent No. 6444799
/ GENERAL INFORMATION:
/ APPLICANT: Ross, Bruce C.
/ TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
/ NUMBER OF SEQUENCES: 1120
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORRISON & FOERSTER
/ STREET: 755 PAGE MILL ROAD
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304-1018
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ OPERATING SYSTEM: IBM Compatible
/ SOFTWARE: FASTSEQ for Windows Version 2.0b
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/221,017B
/ FILING DATE: 23-DEC-1998
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PP1182
/ FILING DATE: 31-DEC-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PP1546
/ FILING DATE: 30-JAN-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PP2911
/ FILING DATE: 09-APR-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/AU98/01023
/ FILING DATE: 10-DEC-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Monroy, Gladys H
/ REGISTRATION NUMBER: 32,430
/ REFERENCE/DOCKET NUMBER: 27340-20021.00
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-813-5600
/ TELEFAX: 650-494-0792
/ TELEX: 706141
/ INFORMATION FOR SEQ ID NO: 646:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 657 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: circular
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: UNKNOWN
/ ORIGIN: SOURCE:
/ ORGANISM: PORPHYROMONAS GINGIVALIS
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 1...657
/ US-09-221-017B-646

Alignment Scores:
Pred. No.: 2.95e-28 Length: 657
Score: 328.50 Matches: 71
Percent Similarity: 55.26% Conservative: 34

```

Best Local Similarity: 37.37% Mismatches: 78  
 Query Match: 6.99% Indels: 7  
 DB: 4 Gaps: 3

US-10-070-464-1 (1-882) x US-09-221-017B-646 (1-657)

```

QY 614 TTTTCTTGTGATGTTTCACTATCAAACTCAATCGGCTTTGAATGATGCTCG 633
DB 566 TTAACCCGAGAGGTTTACACTATCAAACTCAATCGGCTTTGAATGATGCTCG 507
QY 634 LeuTyrLysProHisAspLeuGlnProGlyLysTyrProThrValLeuPheIleTyr 653
DB 506 ATCGGTAACCTATTGATTTCCATCCCTCCGCACTATCTGCTGATGTAAGTAT 447
QY 654 GJYGLYPTGGLNValGlnLeuValAaAnaMaArgPheLysGlyValLysTyrPheArgLeu 673
DB 446 AGCGGCTCCAACTCCAGAGGATTTGATCGCTAT-----TCATTGATGGGA 396
QY 674 AenThrLeuAlaSerLeuGlyTyrValValValIleAspAnaArgLysSerCyHis 693
DB 395 CACTACCTTGATCGAAGGTTACGTGCGCATGTGGATGGCGTGGCAACGGTGTCT 336
QY 694 ArgGlyLeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAsp 713
DB 335 CGCGGCGAAGATGCGCAAGTGTACTCATGCAATGCACTGCTGATTCGAAAGCGATGAT 276
QY 714 GlnValGlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheLeuAspArgVal 733
DB 275 CAGATAGCAGCGCCACTGCTATA--GGACAGCTGCCCTATGTGATGATCACTGCTATTC 219
QY 734 GJYLIeHISGLYTPSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSer 753
DB 218 GGCATATGGGGGTGGAGCTATGGCGGCTATACCACTATAGATTTGTGTGGGGGAAT 159
QY 754 AspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspIle 773
DB 158 GGTACATTCMAAGCGGGGATAGCCGCTGCTGCTGCGCACTGGCGTTCCTACGATTCG 99
QY 774 GJYTYrThGluArgTyrMetGlyHisProAspGlnAnaGlnGlyTyrTyrLeuGly 793
DB 98 GTTTACACCGAAGCTTCAATGGCTACCAACCAAGAGAAATGCTTCCGATTAAGAATGTCT 39
QY 794 Ser-----ValAlaMetGlnAlaGln 800
DB 38 TCTGCTTGTGATGTGGCAAGCCAATTACAA 9

```

RESULT 13  
 US-09-221-017B-253  
 ; Sequence 253, Application US/09221017B  
 ; Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: ROSS, Bruce C.  
 TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
 NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER  
 STREET: 755 PAGE MILL ROAD  
 CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU96/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: MONROY, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 253:

SEQUENCE CHARACTERISTICS:

LENGTH: 543 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: FORYPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..543

US-09-221-017B-253

Alignment Scores:

Score: 2,42e-28 Length: 543

Percent Similarity: 328.00 Matches: 63

Best Local Similarity: 58.82% Conservative: 37

Query Match: 6.98% Indels: 66

DB: 4 Gaps: 2

US-10-070-464-1 (1-882) x US-09-221-017B-253 (1-543)

```

QY 676 LeuAlaSerLeuGlyTyrValValValIleAspAnaArgLysSerCyHisArgGly 695
DB 43 ATGGCAGCAAGAGGCTATGCCGCTTTACGGGATAGCGCGGATCTCCATTAAGAGG 102
QY 696 LeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspArgVal 715
DB 103 GCTGCTTTCAGCAGGTTATTATGTCGTTGGGGCAGACCGAGATGCCGATCAAGATG 162
QY 716 GlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 735
DB 163 TCGGGTGTGATTTCTCTCAAGACCA--TCATGGTGGAATGCCGATAGATAGAGTA 219
QY 736 HisGlyTyrPserTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 755
DB 220 CATGGCTGGAGCTATGTGGCTTTATGACTACGAATCTGATCTTAACGACCGCGATGTC 279
QY 756 PheArgValAlaIleAlaGlyValaProValThrLeuTyrIlePheTyrAspThrGlyTyr 775
DB 280 TTCAAGTCGAGTATGCGCGGCTGTGCAATGATGAATGATATGATATGATGATG 339
QY 776 ThrGluArgTyrMetGlyHisProAspGlnAnaGlnGlyTyrTyrLeuGlySerVal 795
DB 340 GGTGAGCGTTATTTGATGCGCCACAGAAATCCGAAAGATACGATGCTGCCAACCTG 399
QY 796 AlaMetGlnAlaGlyLysPheProSerGluProAnaArgLeuLeuLeuHisGlyPhe 815
DB 400 CTCAAACGAGCCGGATCTG-----AAAGACGACTTATGATGATGAGCG 450
QY 816 LeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAla 835

```

Db	451	ATCGATCCG	GTG	TGG	TAT	GGA	GGA	TTT	ACT	CTTT	CTT	GAT	GCTT	GCG	TAA	GCA	510
Qy	836	Gly	Leu	Pro	Tyr	Asp	Leu	Gln	Ile	Tyr	Pro						845
Db	511	CGCACTTAT	CTG	ACT	CTT	TAC	GCT	TAT	CCG								540

## RESULT 14

```

US-09-016-080-2
; Sequence 2, Application US/09016080
; Patent No. 613012
;
GENERAL INFORMATION:
APPLICANT: Ishikawa, Kazuhiko
APPLICANT: Matsu, Ikuo
APPLICANT: Ishida, Hiroyasu
APPLICANT: Kosugi, Yoshisugu
APPLICANT: Higuchi, Katsuhiko
TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
FILE REFERENCE: 07898/022001
CURRENT APPLICATION NUMBER: US/09/016.080
CURRENT FILING DATE: 1998-01-30
EARLIER APPLICATION NUMBER: JAPAN 18361/1997
EARLIER FILING DATE: 1997-01-31
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.0
;
SEQ ID NO 2
;
LENGTH: 1896
;
TYPE: DNA
;
ORGANISM: Pyrococcus horikoshii
;
FEATURE:
;
NAME/KEY: CDS
;
LOCATION: (1)..(1896)
;
US-09-016-080-2

```

Alignment Scores:	
Pred. No.:	4,386-22
Score:	282.00
Percent Similarity:	28.28%
Best Local Similarity:	23.89%
Query Match:	6.00%
DB:	3
	Gaps: 33
	Matches: 151
	Mismatches: 242
	Indels: 148
	Length: 1899

US-10-070-464-1 (1-882) x US-09-016-080-2 (1-1896)

QY	27	THProSerGIYGLYLySIleuHrgrIleuVTrIGuIuVaMaApbIuserGIuVal	255
Db	199	TCTCCAAATGGGAAGCTTATACGATTTTACCTCAAAAGGGATTAAGAAAGGAAAGAAATCA	258
QY	296	GIuIleIleHISValThISerProMetLeuGIuThrArgaGAlaApSerPheArgTyr	315
Db	259	GAGCTCTACGTATGCCAGCGATGGGGGAGGGCCAGCTTTTAGCAAAAGTTCAATTAC	318
QY	316	ProLySThGIYThrAlaAsnProLySVAlThrPhelySMeTSerGIuIIeMeT	333
Db	319	GGGATTAAGAACTGGCGTTTACCGAGATGGGAAAGATATAGCCGTGTT	369
QY	334	IIeASpAlaGIuGIYArgIleIleASpValIIeASpLySGuIuLeuIleGIuPro	351
Db	370	ACCCCTATATGAGCTTGAGAAAAAAGGAAATGAT-----GACGTTCACATTATTAAGGAA	423
QY	352	PheGIuIleIleuPheGIuGIYValGIuTyrIIleAlaArgAlaGIYTrPthProGIuGIY	371
Db	424	ATACCATCTCGTTTAAATGGAATT-----GGCTGGATC---TACGGA	462
QY	372	LyYrYrAlaTrpSerIIleuIleuAsp-----ArgSerGIuThrArgLeu-----	386
Db	463	AAAAAGAAAGTGTCTTACCTTGTTATGACGTGAGAGCCGGAGAAAAAGAGCTTAACCTCA	522
QY	387	GIuIleValIleuIleSerProGIuIleuPheIIleProValGIu	400
Db	523	AAGAACCTAAATGTTGATCAGATTAAGGTTCCACAAAGCTAGACTATATCTTCAAGGCCAA	582
QY	401	AspAsp-----ValMetGIuArgIn-----	407

Db	583	GAGGATAGGGAAGGAAACCTCTGATATCCATCTTAAACGTCCTGAGAAATGAAAGTT	642
Qy	408	-----ArgLeuileGuseValProAspSerValThrPro	419
Db	643	AGGAAGCTGACCCACGAGAGTGGAGATATCCAGCTTCCTCCCTTGATAGAGGAAC	702
Qy	420	LeuileileTyrgLugLuthThAsp-----lleTrpIleAenlleHisaplleHe	437
Db	703	TTCCGTACTTAAGGCTAACACTTTAGAAAGGGAAATCCCAACCAAGCCAC---ATCTAC	759
Qy	438	HisValPheProGInserHisGluGluGluIuleGluPheIlePheAlaSerIuCyus	457
Db	760	CACATACGATCCC-----AAG	774
Qy	458	ThrgIlePheArgHisLeuTyrlYslelleThrSerlleLeuLysGluSerLysTyrlYsArg	477
Db	775	ACAGGA-----GAACCTTAAGACGTACCAAGAGATTTAAGACAGAAACGCTTACAAATCC	828
Qy	478	SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla	497
Db	829	TTAAAC-----TCCGATTTGGAGGAAGTCAAGAGCCGACCTTG	870
Qy	498	lleThrSerGlyLugLuthTrpGluValleuClYlArgHisGlySerAenlleGlnValAspGlu	517
Db	871	TACAAAGAGGG---TGG-----	885
Qy	518	ValArgArgLeuValTyrlPheGluGlyThrlYsAspSerProLeuGlnHisleuTyrl	537
Db	886	-----ACTACTATGTGGCAAG---GATGCCCT---AGGGCAAACTCTTT	927
Qy	538	ValValSerTyrlYsAlaenProGlyLugLuthThArgleuthr-----AspArgGlyTyrl	555
Db	928	AGGGTCAACTTA-----GATGGAAGATTCAAAAGGGTAAGTGAAGTGAAGAACCGTT	981
Qy	558	SerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrlSerAsnGln	575
Db	982	GAA-----AGCTTCATATAGGGGATTAACATCACTTTC	1011
Qy	576	LysAsnProHisCysValSerleuTyrlYsleuSerSerProGluAspAspProThrlCys	595
Db	1015	ACGAGCTCAAGATGCTGTAACCCCAACGAGCTGTATCATATACAGGATGGAAGAGAAAG	1074
Qy	596	LysThrlYsGluPhe-----TrpAlaThrIleLeuAspSerAlaGlyProleuProAsp	613
Db	1075	AAGGTTACCGACTTAAACAAATGG-----ATAAAGGCT	1107
Qy	614	TyrThr-----ProProGluIlePheSerPheGluSerThrThrgIlePheThrlLeuTyrl	631
Db	1108	TACACCTTTCAAAACCTGAACACTTTAAGTTAAAGCAAGTACGGGGGTTGAATATAGT	1166
Qy	632	GlyMetLeuTyrlYsProHisAspLeuGlnProGlyLysLysTyrlProThValleuPhe	651
Db	1168	GCTGGGGTAATGAAGACCGGTGAACCTTCAGAAAGAAAGAAAGATTCACGATTTCTAAG	1222
Qy	652	lleTyrlYsGlyProGlnValGlnleuValAsnAsnArgPheLysGlyValLysTyrlPhe	671
Db	1228	ATTCACCGGTGGTCTTAAACCGCTTAC-----GGTTACGCTTTTATG	1268
Qy	672	---ArgLeuAenThrLeuAlaSerLeuGlyTyrlYsValValValAlleAspAsnArgGly	690
Db	1270	CAGAGTTTCAACGTTTAACTCTTAAGGCTTCTGCTGTGATATTTCTCAATCTTCAAGGG	1328
Qy	691	SerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrlYsMetGlyGlnlleGlu	710
Db	1330	AGCATGTGCTAACGAGAGAGATTC---GCGGATTTAAGGGGACACTTGGGGAGAGGAT	1386
Qy	711	lleAspArgGlnValGluGlyLeuGlnTyrlLeuAlaSerArgTyrlAspPheIleAspLeu	730
Db	1387	TACCAAGATTTAAAGAGAGTGTGATGATGAAGCATTAAGAGATTTGACTCATGATGAGG	1446
Qy	731	AspArgValGlylleHisGlyTrpSerTyrlGlyTyrlLeuSerLeuMetAlaLeuMet	750





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 16, 2003, 00:05:32 ; Search time 659.423 Seconds

(without alignments)  
3512.533 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAMETEDLGVEIFETADC.....HLHYIQLNIGSRIALKVI 882

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Delop 6.0 , Delext 7.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh

-Q=/cgnt2\_1/USPTO.spool/US10070464/runat\_15102003\_113555\_24924/app.query.fasta\_1.2652

-DB=published Applications NA -Qfmt=fastap -SUFIX=rmdb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bl0sum62

-TRANS-human40.ccl -LIST=45 -DOCALLIGN=200 -THR SCORE=ptc -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US10070464@CGN.1.1.845@runat\_15102003\_113555\_24924

-NCPU=6 -ICPU=3 -NO MAP -LARGESQERY -NEG\_SCORES=0 -WAIT -DSRBLOCK=100 -24924

-FANGLOG -DEV.TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:

1: /cgnt2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgnt2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgnt2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgnt2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgnt2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgnt2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgnt2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgnt2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*

9: /cgnt2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*

10: /cgnt2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*

11: /cgnt2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*

12: /cgnt2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*

13: /cgnt2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*

14: /cgnt2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*

15: /cgnt2\_6/prodata/2/pubpna/US10A\_NEW\_PUB.seq:\*

16: /cgnt2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*

17: /cgnt2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4700	100.0	2649	12	US-10-054-776-1

2	4700	100.0	2649	12	US-10-170-789-39	Sequence 39, Appl1
3	4700	100.0	2671	10	US-09-976-674-2	Sequence 2, Appl1
4	4700	100.0	3143	12	US-10-170-789-37	Sequence 37, Appl1
5	4680	99.6	4829	10	US-09-976-674-12	Sequence 12, Appl1
6	4385.5	93.3	4685	10	US-09-976-674-22	Sequence 22, Appl1
7	4385	93.1	4676	10	US-09-976-674-20	Sequence 20, Appl1
8	4092.5	77.9	4523	10	US-09-976-674-8	Sequence 8, Appl1
9	3661.5	77.9	4309	10	US-09-976-674-14	Sequence 14, Appl1
10	2870	61.1	2617	10	US-09-976-674-4	Sequence 4, Appl1
11	2870	61.1	4219	10	US-09-976-674-28	Sequence 28, Appl1
12	2870	61.1	4302	10	US-09-976-674-24	Sequence 24, Appl1
13	2820.5	60.0	4180	10	US-09-976-674-36	Sequence 36, Appl1
14	2820.5	60.0	4263	10	US-09-976-674-34	Sequence 34, Appl1
15	2649	56.4	4076	10	US-09-976-674-32	Sequence 32, Appl1
16	2649	56.4	4159	10	US-09-976-674-30	Sequence 30, Appl1
17	2638	56.1	2801	13	US-10-098-841-100	Sequence 100, Appl
18	2599.5	55.3	4037	10	US-09-976-674-40	Sequence 40, Appl1
19	2599.5	55.3	4120	10	US-09-976-674-38	Sequence 38, Appl1
20	2476.5	52.7	3262	13	US-10-098-841-83	Sequence 83, Appl1
21	1400	29.8	2079	13	US-10-044-090-843	Sequence 843, App
22	1391	29.6	2411	10	US-09-976-674-26	Sequence 26, Appl
23	1278	27.2	1356	10	US-09-976-674-10	Sequence 10, Appl
24	1026.5	21.8	832	10	US-09-976-674-18	Sequence 18, Appl
25	805	17.1	502	11	US-09-918-995-19585	Sequence 19585, A
26	665.5	14.2	561	11	US-09-764-891-877	Sequence 877, App
27	529	11.3	3407	12	US-10-423-714-5	Sequence 5, Appl1
28	529	11.3	3407	13	US-10-002-593-5	Sequence 5, Appl1
29	529	11.3	3407	14	US-10-165-603-6	Sequence 6, Appl1
30	524	11.1	4835	10	US-09-917-800A-1570	Sequence 1570, Ap
31	524	11.1	4835	14	US-10-165-603-5	Sequence 5, Appl1
32	505	10.7	2130	14	US-10-156-761-3131	Sequence 1131, Ap
33	505	10.7	9025608	14	US-10-156-761-1	Sequence 1, Appl1
34	464	9.9	620	10	US-09-976-674-16	Sequence 16, Appl1
35	461	9.8	2388	12	US-09-870-133-3	Sequence 3, Appl1
36	461	9.8	2388	14	US-10-160-501-6	Sequence 6, Appl1
37	461	9.8	2583	10	US-09-976-674-6	Sequence 6, Appl1
38	461	9.8	3238	12	US-09-870-133-1	Sequence 1, Appl1
39	461	9.8	3238	14	US-10-160-501-4	Sequence 4, Appl1
40	461	9.8	4541	10	US-09-976-674-42	Sequence 42, Appl
41	456	9.7	2366	12	US-10-101-510-683	Sequence 683, App
42	456	9.7	2788	12	US-10-269-909-22	Sequence 22, Appl
43	456	9.7	3128	14	US-10-198-845-13171	Sequence 13171, A
44	455	9.7	3224	12	US-10-240-965-117	Sequence 117, App
45	454	9.7	2814	10	US-09-962-832-108	Sequence 108, App

## ALIGNMENTS

RESULT 1  
US-10-054-776-1  
; Sequence 1, Application US/10054776  
; Publication No. US20030165818A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark Robert Edbrooke  
; APPLICANT: Alan Peter Lewis  
; TITLE OF INVENTION: NOVEL PROTEIN  
; FILE REFERENCE: OG1042US  
; CURRENT APPLICATION NUMBER: US/10/054, 776  
; CURRENT FILING DATE: 2002-01-23  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2649  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: CDS  
; LOCATION: (1)..(2649)  
US-10-054-776-1  
Alignment Scores: 0  
Pred. No.: 4700.00  
Length: 2649  
Matches: 882

Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 12  
Gaps: 0

US-10-070-464-1 (1-882) x US-10-054-776-1 (1-2649)

QY 1 MetAlaAlaMetGluThrGluGluLeuGlyValGluIlePheGluThrAlaAspCys 20  
Db 1 ATGCACACAGCAATGGAAACAGAAACAGCTGGGTGTGGATATTTGAAATCGCGACGT 60  
QY 21 GluGluAsnIleGluSerGluAspAspProlyserGluProPheTyValGluArgTyr 40  
Db 61 GAGAGAAATATTGATATCACAGATCGGCTTAAATTGGAGCCTTTTATGTTGACGGTAT 120  
QY 41 SerTrpSerGluLeuLysLysLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60  
Db 121 TCCTGGAGTCACTTAAAGCTGCTTCCGATACAGAAATATCATCGCTACATGATG 180  
QY 61 AlAlaValProHisAspPheMetCysValLysArgAsnAspProAspGlyProHisSer 80  
Db 181 GCTAAGGACCACTGATTCATGTTGTGAAGAGATGATCCAGATGACCTCATTTCA 240  
QY 81 AsparGlyIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100  
Db 241 GACGAAATCTTATACCTTGGCCATGCTGTGTGAGAACAGAAATATCATCTTTTATTTCT 300  
QY 101 GluIleProlystrIleAsnArgAlaAlaValIleuMetLeuSerTrpLysProLeuLeu 120  
Db 301 GAAATTCGCAAAACATCATATAGACAGACAGCTTAAATGCTCTCTGGAGACCTCTTTTG 360  
QY 121 AspleuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140  
Db 361 GATTTTTTTCGGGACACCTGACTATGGAATGATTTCTCGAAGAGAAACTATTTAGA 420  
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160  
Db 421 GAAAGAAACCATTTGGAAACAGTGGAAATTCCTTTACATTTATCACAAAGAAAGTGGAA 480  
QY 161 ThrPheLeuPheGluAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180  
Db 481 ACATTTCTGTTTCAAGCGGTAGTGAATTTATCATCGTAAAGATGGAGGCGCAAGAA 540  
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
Db 541 TTATCGAACAACCTTTAAGGCCCAATCTAGTGAAGAACTAGTTCTCCCAACATACGATG 600  
QY 201 AspProlyserLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220  
Db 601 GATCAAAATTTATGCTCCGTGATCCAGATCGATGCTTTTATACATAGCAACGATATT 660  
QY 221 TrpIleSerAsnIleValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240  
Db 661 TGGTATCTAACAATCGTAAACAGAGAAAGAGACTCACTTATGTGCAATAGAGCTA 720  
QY 241 AlaAsnMetGluGluAspAlaArgSerArgIleValAlaThrPheValIleuGlnGlu 260  
Db 721 GCCAACAATGGAGAAAGATGCCAGATCAGCTGAGTCTTGTGTTCCAAAGAA 780  
QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280  
Db 781 TTTCATATGATATTCTGGCTATTGGTGTGCTCCAAAGCTGAACATCTCCAGTGGTGT 840  
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGlyValGluIleHisVal 300  
Db 841 AAAATTTCTTGAATTTCTATATGAAGAAATGATGAATCTGAGTGGAAATATTATCATGTT 900  
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
Db 901 ACATCCCTTATGTTGGAAACAGAGAGGAGATTCATTCCTTATCTTAAACAGGTACA 960  
QY 321 AlaAsnProlyserValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340

Db 961 GCAATCTTAAGCTCACTTTTAAGATGTCAGAAATTAATGATTGCTGAAAGAGATC 1020  
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360  
Db 1021 ATGATGTCATAGTAAAGAACTAATTCACACCTTTTGAGATTCATATTGAAAGAGTTGAA 1080  
QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrValAspSerIleLeuLeuAsp 380  
Db 1081 TATATTGCCAGAGCTGATGAGACTCTGAGAGAAATATGCTTGCTCATCTCACTAGAT 1140  
QY 381 ArgSerGlnThrArgLeuGlnIleValIleuIleSerProGluLeuPheIleProValGlu 400  
Db 1141 CGCTCCAGACTCGCTCGAGATGATGTTGATCTCACTGAATTAATTTATCCAGTGA 1200  
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420  
Db 1201 GATGATGTTATGGAAAGCAGAGACTCATTTAGATCACTGCTGATTTCTGTGAGCCACTA 1260  
QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440  
Db 1261 ATTATCTATAGAGAAACAGACATCTGATTAATTTCCATGACATCTTTCATGTTT 1320  
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460  
Db 1321 CCCCAAGTCAAGAGAGAGAAATTTGATTTATTTTCTCTGAAATCAAAACAGTTTC 1380  
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480  
Db 1381 GGTCAATTATACAAATATTAATCATCTTTTAAAGAAACAAATTAACATCCAGTGGT 1440  
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500  
Db 1441 GGGCTGCTCTCCAAATGATTTTCMACTGTCCTATCAAAAGAGATAGCAATTAACAGT 1500  
QY 501 GlyIleTrpGluValIleuGlyArgHisSerAsnIleGlnValAspGluValArgArg 520  
Db 1501 GGTGAATGGGAAGTCTTGGCCGAGATCGATCTTATATCCAAAGTTGATGAAGTGAAGG 1560  
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeuTyrValIleSer 540  
Db 1561 CTGGTATATTTTGAAGGACCAAAAGACTCCCTTTAGAGATCACTGTACGTAGTCACT 1620  
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560  
Db 1621 TACGTAATCTCGAGAGGTGACAGAGGCTGACGAGCCGTGCTACCTACATCTTCTGCTGC 1680  
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580  
Db 1681 ATCAGTACGACATGTGACTCTTTATTAAGTATGATATGTAACCAAGAAATCCACACTGT 1740  
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600  
Db 1741 GTGTCCCTTACAAAGCTATCAAGCTCGAAGATGACCAACTGCAAAACAAAGAAATTT 1800  
QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620  
Db 1801 TGGGCCACATTTTGGATTGACAGAGGCTCTTCTGACATACCTCTCCAGAAATTTTC 1860  
QY 621 SerPheGluSerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640  
Db 1861 TCTTTTGAAGTACTACTGATTTTACCTTTATGAGATGCTTCAACAGCCTCATGATCTTA 1920  
QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660  
Db 1921 CAGCTCGAAGAAATATCTCACTGAGCTGCTCATATATATGCTGCTCAGGTGCACTTG 1980  
QY 661 ValAsnAspArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680  
Db 1981 GTGAATATGAGTTTAAAGAGTCAGATATTTCCGCTGAATATCCCTAGCCTCTTAAGGT 2040  
QY 681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLysLysPheGluGly 700  
Db 2041 TATGTGTTGTATGATAGCAACAGGGATCTGTACCGAGGCTTAAATTTGAAAGC 2100

```

QY 701 AlAphelysTyrrylusMetGlyGlnIleGlnIleAspAspGlnValGluGlyLeuGlnTyr 720
Db 2101 GCGTTTAAATATATAATGGGTCAATATGAAATTACATACAGTGGAAAGATCTCCAAATAT 2160
QY 721 LeuAlaserArgTyrrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPseTyr 740
Db 2161 CTAGGCTCTCGATATGATTTTCATTGACTTAGATGCTGTGGGCATCCAGCGGTGCTCTAT 2220
QY 741 GlyGlyTyrrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db 2221 GGAGGAATACCTCTCCGATGGCATTAATGACAGAGTCAGATATCTTCAGGGTTGCTATT 2280
QY 761 AlAGlyAlaProValThrLeuTyrIlePheTyrrAspThrGlyTyrrGluArgTyrMet 780
Db 2281 GCTGGGGCCCGACGACCTGTGGATCTTCTATATACAGATACAGGAAACGTTATATG 2340
QY 781 GlyHisProAspGlnAsnGlnGlnGlyTyrrLeuGlySerValAlaMetGlnAlaGlu 800
Db 2341 GGTCAACCCCTGACGAATGAACAGGGCATATTACTTAGATCTGTGGCATGCAAGCAGAA 2400
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db 2401 AAGTTCCCTCTGAAACCAATCGTTACTGCTTACATGGTTTCCGAGATGAGAAATGTC 2460
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db 2461 CATTTTGACATACCACTATATTTACTAGTTTTCATGAGGCTGGAAACCCATATGAT 2520
QY 841 LeuGlnIleTyrrProGlnGluArgHisSerIleArgValProGlnSerGlyGlnHisTyr 860
Db 2521 TTACAGATCTATCTCTGAGAGACACAGCATTAAGATCTCGTAATCTGGAGAACATTAT 2580
QY 861 GluLeuHisLeuLeuHisTyrTyrrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys 880
Db 2581 GAACGTGATCTTTGGACTACTGCTTCAAGAAACCTTGATACGTAATGCTGCTTAAAA 2640
QY 881 ValIle 882
Db 2641 GTGATA 2646

```

```

; PRIOR APPLICATION NUMBER: US 60/226,740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-170-789-39

Alignment Scores:
Pred. No.: 0 Length: 2649
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 12

US-10-070-464-1 (1-882) x US-10-170-789-39 (1-2649)
QY 1 MetAlaAlaIleMetGluThrGlnGlnIleGlnIlePheGluThrAlaAspCys 20
Db 1 ATGGACGACCAATGAGAAACAGACAGCTGGCTTGACATATTTGAAATCGCGACTGT 60
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
Db 64 GAGAGAAATATTAATCAACAGATCGGCCCTAAATTGGAGCTTTTATATGTTGAGCGGTAT 120
QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 121 TCTTGAAGTCAGCTTAAAGAGCTGTCGATGACAGAAATATATATGCTACATGATG 180
QY 61 AlaIleAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 181 GCTAAGGACCAATATGATTTTCATGTTTGGAAGAGATATCAAGATGACCTCATTTCA 240
QY 81 AspArgIleTyrTyrrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db 241 GACAGATCTATTAATCTGCTGCTGATGCTGGTGAAGAACAGAGAAATATCACTGTTTATTC 300

```

QY	101	GIUIPEPOLYETHIRLEASNAAGLAALAVALLEUWETLEUSERTRIPLYSEPROLEUEN	120
DB	301	GAANATCCCAAAACATCATATGAGACACAGCTCTTAAGCTCTCTGGAAACCTCTTTTG	360
QY	121	ASPLEUPHEGINALATHLEUASBPYRGIMETYSRSATRGUGUGULEUENRG	140
DB	361	GATCTTTTTCAGGCACACCTGACCTATGGAATGTRTTCTCGAAGAAAGAACTTATAGA	420
QY	141	GIUARGYLSARGILEGYTHRVALGYILEASERYRAEPTYRHISGLINGLYSERGLY	160
DB	421	GAAGAAACCGATTCGTAAGAACAGTCGGAATTCCTCTTCAGATTATACCAAGAAAGTGA	480
QY	161	THRPHELEUPHEGINALAGLYSERGLYILEYTRHISVALVLSAEPGLYGLYPROGLYGLY	180
DB	481	ACATTTCGTTCAGAGCCGGTATGTAATTTATCACGTAAGAAAGAGGGCCACACAGGA	540
QY	181	PHETHRINGINPROLEUARGPROMSULEUVALGIUTHSERCYSPROASNILEARGMET	200
DB	541	TTTAGCAACAACCTTTAAGGCCCAATCTAGTGAAGAACTAGTGTGCCAACATACGAGTG	600
QY	201	ASPPOLYBLEUCYSPROALASBPROMSPTPIELALPHEILEHISERASNAAPILE	220
DB	601	GATCCAAATAATATGCCCGCTGATCCAACTGGATTCGTTTATATACATAGCAACAGATATT	660
QY	221	TRIPLESERASNILEVALTHARGGLUGIUAARGYLEUTHRYTRHYVALHISASNGIULEU	240
DB	661	TGGATATCTTAACATCTGTAACAGAAAGAAAGAAAGACCTCATTTATGTGCACAAATGAGCTA	720
QY	241	ALASENMETGIUGIUAASPALAARGSERALAGLYVALATHRPHIVALILEUGIUGIUGU	260
DB	721	GCCAAACATGGAAGAAAGATGCCGAGTCACTGGAGTCGCTACCTTGTGTTCTCCAAAGAA	780
QY	261	PHASPAARGYRSEGLYTRYPTRIPCYSPOLYVALAGIUTRTHRPROSEGLYGLY	280
DB	781	TTGTGTAATATATCTGGCTATTGTGGTGTCACAAAGCTGAAACACTCCCAAGTGGTGT	840
QY	281	LYSILELEUAAGILELEUYRGUGIUNASNAAPGUSERGUVALGIUIELEHISVAL	300
DB	841	AAAATTTCTAGATTCTATATGAAGAAATGATGAATCTGAGGTGGAATATTATCATGTT	900
QY	301	THRSERPROMETLEUGIUTHRARGARGLAAPSPSERPHEAARGYRPROLYSETHGYLTHR	320
DB	901	ACATTCCTCTATGTGGGAAACAAAGGGCCAGATTCATTCGGTATCTCTMAAACAGATACA	960
QY	321	ALASNPOLYVALTHRPHELYMETSERGUILEMETILEAPALAGIUGLYARGILE	340
DB	961	GCAAACTCTTAAGTACCTTTTAAGATGTGCAGAAATATGATGATGCTGGAAGGAAGATC	1020
QY	341	ILEASPVALLLEAPLYGULEUULEGINPROHEGIUIELEUPHEGIUYVALIGU	360
DB	1021	ATAGATGTCAATAGATMAAGAACTAATTCAACCTTTTGATTTCTAATTTGAAGAGGATGGA	1080
QY	361	TYRILEALATGALAGLYTRIPTHRPROGLUGIULYSTRALATTPSERILELEUENASP	380
DB	1081	TATATGCCAGAGCTGGATGACATCTCTAGGAAATAATGCTGTGCTCATCTACTAGAT	1140
QY	381	ARGSERGINTHRARGLEUINILEVALLEUIESERPROGIULEUPHEILEPROVALIGU	400
DB	1141	CGCTCCAGAGCTCGGCTCGAGTATGTTGATCTCACCTGAATATTATTATCCAGATGAA	1200
QY	401	ASPAAPVALMETGIUARGIUMARGLEUULEGIUSERVALPROAPSPSERVALTHPROLEU	420
DB	1201	GATGATGTTATAGAAAGGACAGATCTCTTAAGTCAAGGCTGTGATTCGTAGCCCACTA	1260
QY	421	ILEILEYRGUGIUTHRTHASPILETPIRILEASNILEHISAPPILEPHEHISVALPHE	440
DB	1261	ATTATCTATGAAGAAACAACAGACATCTCGAATAATATCATGATCTTTCACTGTTT	1320
QY	441	PROGINSERHISGLUGIUGIUILEGIUPHEILEPHEALASERGIUCYLYSETHGYLPH	460
DB	1321	CCCCAAATCAACGAAGAGAAATGACGTTTATTTTGGCTCTGATAGCAAAACAGGTTTC	1380

QY	461	ArgHisbLeuTYrVleIleThSerIleLeuLysgluSerLysTYrLysArgSerSerLys	480
Db	1381	CGTCATTATATCAAAATATACATTCATATTTAAAGAAAGCAAAATATATAACGATCCGCTGT	1440
QY	481	GlyLeuProAlaProSerAspPheLysGlySerProIleLysgluGluIleAlaIlePheSer	500
Db	1441	GGCGCTGGCTGCCAAGATTTCAAGTTCCTATCAAAAGGAGATAGCAATTAACAGT	1500
QY	501	GlyLysTrpGluValLeuGlyArgHisGQYSerAsnIleGluValAspGluValArgArg	520
Db	1501	GGTGAATGGAAAGTTCTTGCGCGGACATGAGCTATATACCAAGTTGATGAAGTCAGAAAG	1560
QY	521	LeuValTYrPhegluGlyThrLysAspSerProLeuGluHisbLeuTYrValValSer	540
Db	1561	CTGGATATATTTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGATGATGTCAGT	1620
QY	541	TyrValAsnProGlyLysGluValThrArgLeuThrAspArgGlyTYrSerHisSerCysGys	560
Db	1621	TACGTAAATCCTGGAGAGGTACAAAGGCTGACTACCGTGGCTACTCATATTCCTGCTGC	1680
QY	561	IleSerGlnHisCysAspPhePheIleSerLysTYrSerAsnGlnLysAsnProHisbCys	580
Db	1681	ATCAGTCACGACTGTGACTTCTTATATAGTATAGTATACCAAGAAATCCACACTGT	1740
QY	581	ValSerLeuTYrLysLeuSerSerProGluAspAspProThrCysLysThrLysgluPhe	600
Db	1741	GTCGCCCTTTCAAGCATATCAAGCTCGTAGATGATCCCAACTTCCAAACCAAGAAATTT	1800
QY	601	TrpIleThrIleLeuAspSerAlaGlyProLeuProAspTYrThrProProGluIlePhe	620
Db	1801	TGGGCACACATTTGGATTGATTCAGCAGGTCTCTTCTGATCATATCTCTCCAGAAATTTTC	1860
QY	621	SerPhegluSerThrThnGlyPhePheThrLeuTYrGlyMetLeuTYrLysProHisbAspLeu	640
Db	1861	TCTTTGAAATACTACTGAGATTACATTCATATGATGGATGCTCTACCAAGCTCATGATTTA	1920
QY	641	GlnProGlyLysLysTYrProThrValLeuPheIleTYrGlyProGlnValGlnLeu	660
Db	1921	CAGCTGGAAAGAAATATCTCACTGCTGCTTCAATATATGGTGGTCCAGGTGCAGTTG	1980
QY	661	ValAsnAsnAspPheLysGlyValLysTYrPheArgLeuAsnThrLeuAlaSerLeuGly	680
Db	1981	GTCGAATATCCGTTTAAAGAGTCAAGATTTCCGCTTGAATACCTTAGCTCTCTAGCT	2040
QY	681	TYrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly	700
Db	2041	TATGTGCTTGATAGCATAGACACAAAGGGATCCTGTCAACGAGGCTTAAATTTGAAGGC	2100
QY	701	AlaPheLysTYrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTYr	720
Db	2101	GCCTTTAAATATATAAATGGGTCAAATATGAAATTCATGACGATCGGTGGAAGGACTCCAATAT	2160
QY	721	LeuAlaSerArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTYrPheTYr	740
Db	2161	CTACCTCTCGATATGATTCATTCATTGACTTCAGTATGATGTGTGGGCATCAACGGCTGTGCTAT	2220
QY	741	GlyGlyTYrLysLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle	760
Db	2221	GGAGAGATACCTCTCTCCGATGACCTTAATATGACAGAGGTACAGATATCTTCAGGGCTGTATTT	2280
QY	761	AlaGlyAlaProValThrLeuTrpIlePheTYrAspThrGlyTYrThrGluArgTYrMet	780
Db	2281	GCTGGGGGCCAGTCACCTGTGTGATCTTCTATGATACAGGATACAGGAACGTTATATG	2340
QY	781	GlyHisbProAspGlnAsnGluGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGlu	800
Db	2341	GGTCAACCTTGACCAATATGAAACAGGCTATTTACTTAGATCTGTGGCCATGACAGAGAA	2400
QY	801	LysPheProSerArgLysProAsnArgLeuLeuLeuHisbGlyPheLeuAspGluValAsnVal	820
Db	2401	AAGTTTCCCTCTGACCAATCGTTTACGCTCTTACATGATGGTTCCTGAGATGAGAAATGTC	2460
QY	821	HisbPheAlaHisbThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAsp	840

Db 2461 CATTTGCACATACAGATATTAAGTCTTTTGTAGTGAAGGCTGGAAACCATATGAT 2520  
Qy 841 LeuGlnIleTyrrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860  
Db 2521 TTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATCTCTGATCGCGGAACATTAT 2580  
Qy 861 GluLeuHisLeuLeuHisTyrLeuGlnGluLeuGlnGlySerArgIleAlaIleLeuHis 880  
Db 2581 GAATGCAATCTTTTGCACTACCTTCAGAAAACTTGGATCAGTATTGCTGCTTAAAA 2640  
Qy 881 ValIle 882  
Db 2641 GTGATA 2646

RESULT 3  
US-09-976-674-2  
; Sequence 2, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akimsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2671  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-976-674-2

Alignment Scores:  
Pred. No.: 0 Length: 2671  
Score: 4700.00 Matches: 882  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-070-464-1 (1-882) x US-09-976-674-2 (1-2671)

Qy 1 MetAlaAlaIaIaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20  
Db 8 ATGGCAGCAGCAAGAAAGAAACAGACAGCTGGGTCTTGAGATATTGAAACTGCCGACTGT 67  
Qy 21 GluGluAsnIleGluSerGlnAspArgProIleGluGlnProPheTyrValGluArgTyr 40  
Db 68 GAGGAGATATTGATATACAGAGATCGGCTTAATTGAGCTTTTATGTGACCGGTAT 127  
Qy 41 SetTrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60  
Db 128 TCCCGAGGTGAGCTTAAGAAAGCTGCTTGCCGATACCGAAATATCATGCTACATATG 187  
Qy 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80  
Db 188 GCTTAAGCACAACATGATTTTCATGTTGTGAAGAGATATCCAGATGGACCTCATTTCA 247  
Qy 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgLysLeuThrLeuPheTyrSer 100  
Db 248 GACAGATCTATTAACCTTGCCATGCTCGTGAAGAACAGAAAAATACACTGTTTATTTCT 307  
Qy 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120  
Db 308 GAATTTCCCAAACTATCAATATAGACAGCACTTTAATGCTCTCTTGGAAGCCCTTTTG 367  
Qy 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140

Db 368 GATCTTTTTCAGGACAACCTGAGATATGATATGATCTGAGAAAGAACTATTAGA 427  
Qy 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160  
Db 428 GAAAGAAAGCAATTTGGAACAGTGGAAATGCTTTCAACGATATATCACCAAGAGAGTGA 487  
Qy 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspArgGlyProGlnGly 180  
Db 488 ACATTTCTGTTCAAGCCGGAGTGAATTTATCATCTTAAGATGAGAGGCCCAAGGA 547  
Qy 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
Db 548 TTTACGCAACACCTTTTAAGCCCAATCTAGTGAACCTAGTTCTCCACATACGATG 607  
Qy 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220  
Db 608 GATCCAAATTTATCCCTGCTGATCCAGACTGATGCTTTTATATACATAGCAACGATATT 667  
Qy 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240  
Db 668 TGGATATCTTAACATCGTACACAGAGAGAGAAAGAGACTCTTATGTGCACAAATGACTTA 727  
Qy 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260  
Db 728 GCCAACATGGAAGAGATGCCAGATACGCTGAGTCCGCTACTTGTCTTCCAAAGAA 787  
Qy 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280  
Db 788 TTTGATAGATATCTGCTATTTGGTGTGTGCCAAAGCTGAACACTCCACAGTGCTGT 847  
Qy 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGlyValGluIleHisVal 300  
Db 848 AAAATTCCTAAGATCTATATGAAAGAAATGATGAATCTAGAGGAAATTAATTCATGTT 907  
Qy 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
Db 908 ACATCCCTATGTTGGAACAGAGGAGGACGATTCATTCCTTAAACAGGTACA 967  
Qy 321 AlaAsnProLysValThrPheLysMetSerGlyIleMetIleAspAlaGluGlyArgIle 340  
Db 968 GCAAAATCTTAAGCTATTTTAAGATGTCAAGAAATATATGATTTGCTGAAGAGAAATC 1027  
Qy 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360  
Db 1028 ATGATGTCTATGATTAAGAACTAATTCACCTTTTGAAGATTTCAATTTGAAGGCTTGA 1087  
Qy 361 TyrIleAlaArgAlaGlyTrpThrProGlnGlyLysTyrValaTrpSerIleLeuLeuAsp 380  
Db 1088 TATTTGCCAGAGCTGATGATGACTCTGAGGAAATATGCTTGCTCATCTTACTAGAT 1147  
Qy 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
Db 1148 CGCTCCAGACTCGCTGAGATAGTGTGATCTCACCTGAATATTATCCACGATGAA 1207  
Qy 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420  
Db 1208 GATGATGTTATGGAAGGCGAGACATCATTAAGTCACTGCTGATCTGAGACGCCACTA 1267  
Qy 421 IleIleTyrGlnGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440  
Db 1268 ATTATCTATTAAGAAACACAGACATCTGATTAATTCATACATCTTTTCATGTTTTT 1327  
Qy 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460  
Db 1328 CCCCAAGTCAAGAAAGGAAATTTGATTTTGGCTGTAATGCAAAACAGGTTTC 1387  
Qy 461 ArgHisLeuTyrTyrIleThrSerIleLeuLysGluSerTyrTyrLysArgSerSerGly 480  
Db 1388 CGTCAATTTATACAAATATATCTATTTTAAAGAAAGCAAAATATTAACATCCAGTGT 1447  
Qy 488 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500

Db 1448 GGGCTGCTGCTCCAAGTATTTCAAGTGCTCCATCAAGAGAGATAGCAATTACAGT 1507  
Qy 501 G1YGLUTPRGLUValLeuG1YarGh1sG1YSerAan11eG1nValAepG1ValArGaG 520  
Db 1508 GGTGAATGGGAAGTCTTGCGCGGACATGATCTAAATATCAAGATGATGAATCGAAGG 1567  
Qy 521 LeuVal1YrPhG1uG1YThrLysAAsPserProLeuG1uH1sH1sLeuTYrVal1A1Ser 540  
Db 1568 CTGGTATATTTTGAAGGACCAAGACTCCCTTTAGACATACCTGTAAGTACAGT 1627  
Qy 541 TYrValAsnProG1YGLUVal1ThrArGLeuThrAspArg1YTYrSerH1sSerCysCys 560  
Db 1628 TACGTAATATCTGGAGAGGTGACAAAGCTGACGACCGTGGCTACTACATTCTGCTCG 1687  
Qy 561 11eSerG1nH1sCysAAsPhePheH1eSerLysTYrSerAAsG1nLysAAsnProH1sCys 580  
Db 1688 ATCAGTCAGACTGAGACTTCTTTATATAGTATAGTATACAGACAGATCCACACTGT 1747  
Qy 581 ValSerLeuTYrLysLeuSerSerProG1uAAsPAsPProThCysLysThrLysG1uPhe 600  
Db 1748 GTGTCTCTTACAGCTATCAAGTCTCTGAAGATGACCAACTTGCAAAACAAAGAAATT 1807  
Qy 601 TTPA1ArH11eLeuAAsPserA1aG1YProLeuProAsPTYrThrProG1u11ePhe 620  
Db 1808 TGGGCAACATTTGGATTCAAGAGGTCTCTTCTGACTATATCTCTCCAGAAATTTTC 1867  
Qy 621 SerPheG1uSerThrThG1YpHeThrLeuTYrG1YMeLeuTYrLysProH1sAsPLeu 640  
Db 1868 TCTTTTGAAGTACTACTGATTTTAAATGATAGGATCTCTACAGCTCATGATCTCA 1927  
Qy 641 G1nProG1YLysLysTYrProThrValLeuPheH1eTYrG1YGLYProG1nValG1nLeu 660  
Db 1928 CAGCTCGAAGAAATATCTCTACTGTGCTTCAATATAGTG1G1GCTCTAGTGCAAGTTG 1987  
Qy 661 ValAsnAsnArgPheLysG1YValLysTYrPheArgLeuAsnThrLeuA1aSerLeuG1Y 680  
Db 1988 GTGAATTAATCGATTTAAAGAGATCAAGTATTTCCCTGTAATACCTCAAGCTCTAGGT 2047  
Qy 681 TYrVal1Val1Val11eAsPAsnArgLysSerCysH1sArgG1YLeuLysPheG1uG1Y 700  
Db 2048 TATGTG1GTG1AGTATGATACAAACAGGGATCTCTGACCGAGGCTTAAATTTGAAGGC 2107  
Qy 701 AlaPheLysTYrLysMetG1YGL1n1eG1u11eAsPAsnG1ValG1uG1YLeuG1nTYr 720  
Db 2108 GCCTTAAATATTAATTAATGGGTCAAAATGAAATTTGAGATCAGTGGAAAGACTCCAAAT 2167  
Qy 721 LeuA1aSerArgTYrAsPheH1eAsPLeuAsPArgValG1Y11eH1sG1YTPSerTYr 740  
Db 2168 CTAGCTTCTCGATATGATATTCATTGACTTACGTCGAGCATCCACGCGTGCCTAT 2227  
Qy 741 G1YGLYTYrLeuSerLeuMetAlaLeuMetG1nArgSerAsn11ePheArgValA1a11e 760  
Db 2228 GAGAGATACCTCTCCCTGATGGCATTAAGCAAGAGGTCAAGATATCTTCAGGGTGTAT 2287  
Qy 761 AlaG1YAlaProVal1ThrLeuTrp11ePheTYrAsPThG1YTYrThrG1uArgTYrMet 780  
Db 2288 GCTGGGGGCCCACTGATCTGTGGATCTTCTATGATACAGGATACCGAAGATTATAG 2347  
Qy 781 G1YH1sProAsPGLnAsnG1uG1nG1YTYrTYrLeuG1YSerValA1aMetG1nA1aG1u 800  
Db 2348 GGTCAACCCGACCAAGTAAACAGGGCTTACTTACGATCTGTGGCCATGCAAGAGAA 2407  
Qy 801 LysPheProSerG1YProAsnArgLeuLeuLeuH1sG1YPhLeuAsnG1uAsnVal 820  
Db 2408 AAGTTCCTCTGAAACCAATATGTTTACTGCTCTTACATAGGTTTCTTGATAGAAATGTC 2467  
Qy 821 H1sPheA1aH1sThSer11eLeuLeuSerPheLeuValA1aArgA1aG1YLysProTYrAsP 840  
Db 2468 CATTTTGACATACACAGTATATTAAGTATTTTATGATAGGGCTGGAAAGCCATATGAT 2527  
Qy 841 LeuG1n11eTYrProG1nLysArgH1sSer11eArgVal1ProG1uSerG1YGLuH1sTYr 860  
Db 2528 TTACAGATCTATCTCAAGAGACACAGCATTAAGATTCGTAATCGGAGAAACATATAT 2587

Qy 861 G1uLeuH1sLeuLeuH1sTYrLeuG1nLysLeuG1uAsnLeuG1YSerArg11eA1a1a1eLys 880  
Db 2588 GAACATGATCTTTGACATACCTTCAGAAACCTTGATCAAGTATGCTGCTCTCAAAA 2647  
Qy 881 Val11e 882  
Db 2648 GTGATA 2653  
RESULT 4  
US-10-170-789-37  
Sequence 37, Application US/10170789  
Publication No. US20030180930A1  
GENERAL INFORMATION:  
APPLICANT: Rachel E. Meyers  
APPLICANT: Olandt, Peter J.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
APPLICANT: Welch, Nadine  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USBS THEREOF  
FILE REFERENCE: 10448-191001  
CURRENT APPLICATION NUMBER: US/10/170,789  
CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797,039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882,166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212,078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934,406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226,740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861,801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19  
PRIOR APPLICATION NUMBER: US 09/801,267  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07138  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,454  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/829,671  
PRIOR FILING DATE: 2001-04-10  
PRIOR APPLICATION NUMBER: PCT/US01/40483  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: US 60/197,508  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/961,721  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: PCT/US01/29904  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: US 60/235,023  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US 10/045,367  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: US 60/246,561  
PRIOR FILING DATE: 2000-11-07  
PRIOR APPLICATION NUMBER: US 09/801,275  
PRIOR FILING DATE: 2001-03-06



```

: PRIOR APPLICATION NUMBER: PCT/US01/07074
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US 60/187,420
: PRIOR FILING DATE: 2000-03-07
: NUMBER OF SEQ ID NOS: 63
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 37
: LENGTH: 3143
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (229) ... (2874)
US-10-170-789-37

```

Qy	1	MetAlaAlaAlaMetCgluThrgluGluLeuGlyValGluIlePheGluThrAlaAspCys	20
Db	229	ATGGCAGCACCAATGGAAAAAGAAACAGCGGGGTGGTGGATATTATTTGAAACGCGGACTGT	288
Qy	21	GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluAspTyr	40
Db	289	GAGAGAAATTGTGAATCACAGGATCGGCTTAATTGGACCTTTTATGCTTGACCGGTAT	348
Qy	41	SerTyrSerGluLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet	60
Db	349	TCTCGGAGTCACGCTTAAAAAGCTGCTGGCTGATACAGAAAAATATCATGGCTACATATAG	408
Qy	61	AlaLysAlaProHisAspPheMetPheValLysValGlyAsnAspProAspGlyProHisSer	80
Db	409	GCTAAGGCAACCAATATTCATGTTTGTGAAGAGAAATGATCCAGATGGACTTCATTTCA	468
Qy	81	AspArgIleTyrTyrLeuAlaMetSerGlyIuAsnArgIuAsnThrLeuPheTyrSer	100
Db	469	GACAGATCTATTACCTTGCGATGTCTGTGTGAAGACAGAAAAATACACTGTTTATCT	528
Qy	101	GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu	120
Db	529	GAAATTCGCCAAACTATCATAGAGACAGACTTAAAGCTCTCTGGAGACCTCTTTTG	588
Qy	121	AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg	140
Db	589	GATCTTTTTCAGGCACACTGACCTATGGAAATGATTTCTCGAAGAAAGAACTATTAAAG	648
Qy	141	GluArgLysArgIleGlyThrValGlyIleAsertyrAspTyrHisGlnGlySerGly	160
Db	649	GAAAGAAAACGCTTGAACAGTCGGAATTCCTTTCAGATTATCACCAAGAAATGGA	708
Qy	161	ThrPheLeuPheGluAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly	180
Db	709	ACATTTCTGTTTCAACCGGTAGTGAATTTATCCGTAAGAAATGGAGGGCCCAAGGA	768
Qy	181	PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet	200
Db	769	TTTACGCAACAACCTTTAAGGCCCAATCTAGTGAATTAAGTTGCTCCACATTCGATG	828
Qy	201	AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle	220
Db	829	GATCCAAAATTATATGCGCTGCTGATCCAGACTGGAATGCTTTTATATACATAGCAAGATAT	888
Qy	221	TyrIleSerAsnIleValThrArgGluGluArgLysLeuThrTyrValHisAsnGluLeu	240
Db	889	TGGATATCTAACCTCTTAACGAGAAAGAAAGAAAGAACTCACTTATGTGCACATAGACTA	948

QY	241	AlaAenMeTgUgluAAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu	260
Db	949	GCMAACATGGAAAGAAATGCCAGATCAGCTGGAGTCGACCTTGTGTTCCAAAGAA	100
QY	261	PheAAspArgTyTyrSerGlyTyrTrpTrpCysPArgValAlaGluThrThrProSerGlyGly	280
Db	1008	TTTGATAGATATTCTTGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCAGTGGGT	106
QY	281	LysIleLeuArgIleLeuTyTgUgluAAspGluSerGluValAlaGluIleIleHisVal	300
Db	1069	AAAAATCTTGAATCTATATGAAABAAATGATGATCTGAGGTGGAATTTATCATGTT	112
QY	301	ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyTyrProLysThrGlyThr	320
Db	1129	ACATCCCTTATGTTGGAAACAAAGGCGACAGTTCATTCCTGTTATCTTAAACAGGTACA	118
QY	321	AlaAenPArgValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle	340
Db	1189	GGAAATCTCTAAAGTACTTTTAAAGTATGTCAGAAATAAATGATGATGCTGAAGAAAGATC	124
QY	341	IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu	360
Db	1249	ATAGATGTCATAGATAAAGAACTAATTACCTTTGAGATTCTATTGAAAGAGTTGAA	130
QY	361	TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyTyrAlaTrpSerIleLeuLeuAsp	380
Db	1309	TATATTGCCAGAGCTGATGGACTCCTGAGGGAATAATGCTGTGATCCTACTAGAT	136
QY	381	ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu	400
Db	1369	CGCTCCAGACTCGCTCAGATGATGTTGATCTCAGCTGAATTTTATCCAGATGAA	142
QY	401	AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu	420
Db	1429	GATGATGTTATGGAAAGCAGACAGACTATTAGTAGAGTCCGATCTGTGACGCCACTA	148
QY	421	IleIleTyTgUgluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe	440
Db	1489	ATTATCTATGAAGAAACAACAGACATCTGGATTAATATCCATGACATCTTTCATGTTTT	154
QY	441	ProGlnSerHisArgUgluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe	460
Db	1549	CCCCAAATCAGAAAGAGAAATGAGTTTATTTTGGCTCGAATGCAAAACAGATTTTC	160
QY	461	ArgHisLeuTyTyrLysIleThrSerIleLeuLysGluSerLysTyTyrLysArgSerGly	480
Db	1609	CGTCAATTATACAAATATTAATCTATTTTAAAGGAACCAATATTAACGATCCAGTGT	166
QY	481	GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleHisSer	500
Db	1669	GGGCTGGCTGCTCCAGATGATTTCAAGTGTCTTATCAAAAGAGATAGCAATTCAGAT	172
QY	501	GlyLeuTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg	520
Db	1729	GGTGATGGGAAAGTCTTGCCCGGCAATGATTAATATCCAAAGTTGATGAAGTCAGAAG	178
QY	521	LeuValTyTyrPheGluGlyTyThrLysAspSerProLeuGluHisIleLeuTyTyrValSer	540
Db	1789	CTGGATATTTTGAAGCACCACAAAGACTCCCTTTAGGATCACTGTAGTACTAGT	184
QY	541	TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyTyrSerHisSerCysCys	560
Db	1849	TACGTAAATCTCGAGAGGTACACAGGCTGACTACCGTGGCTACTCAATTTCTTGCTGC	190
QY	561	IleSerGlnHisCysAspPhePheIleSerLysTyTyrSerAsnGlnLysAsnProHisCys	580
Db	1909	ATCAGTCCGACTGTGACTTCTTATTAAGTAAATAGTATAGTAAACAGAAATCCACACTGT	196
QY	581	ValSerLeuTyTyrLysLeuSerSerProGluAspAspProTrpCysLysThrLysGluPhe	600
Db	1969	GTGTCCTCTTACAAAGCTATCAAGTCCGAAAGATACCACTTGCACAAACAAAGAAATTT	202
QY	601	TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyTyrThrProProGluIlePhe	620



Db 994 TTGTGATAGATATTCTGGCTATTGTGGTGTGTCCTCCAAAAGCTGAAACAATCCCAAGTGTGCT 1053  
 QY LysIleLeuArgIleLeuTyrgIuGluuAspGluSerGluValGluIleIleHisVal 300  
 Db 281 |||||  
 Db 1054 AAAATCTTAGAATCTATATGAAAGAAATGATGATCTGAGTGGAAATTAATTCATGTT 1113  
 QY ThisSerProMetLeuGluThzArgArgAlaAspSerPheArgTyProLeuThzGlyThr 320  
 Db 1114 ACATCCCTATGTTGGAAACAAGAGGGAGATTCAATCCGTTATCCCTAAACAGAGTACA 1173  
 QY 321 AlaAsnProLygValThrPheLyMetSerGluIleMetIleAspAlaGluGlyArgIle 340  
 Db 1174 GCAATTCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATGATGCTGAAGAGAAAGATC 1233  
 QY 341 IleAspValIleAspLyGluLeuLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360  
 Db 1234 ATAGATGCTATGATAGAAAGAACTAATTCACCTTTTGAGATTCTATTTGAGAGAGCTAGA 1293  
 QY 361 TyrIleAlaArgAlaGlyTyrPheProGluGlyLygTyraIatPserIleLeuLeuAsp 380  
 Db 1294 TATATTGCCAGAGCTGAGTGGACTCTCTGAGGAAATATGCTGTGCTCATCTCTACTAGAT 1353  
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400  
 Db 1354 CGCTCCCAAGACTCCGCTACAGATAGTGTGATCTCACCTGAATTAATTAATCCAGTAGAA 1413  
 QY 401 AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu 420  
 Db 1414 GATGATGTTATGGAAGGCAAGAGACTCATGAGTACAGTCCGCTGATCTGTGAGCGCACATA 1473  
 QY 421 IleIleTyrgIuGluThzThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440  
 Db 1474 ATTAATCTATGAAAGAAACAAGACATCTGATTAATATCCATGACATCTTTCACTGTTTT 1533  
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThzGlyPhe 460  
 Db 1534 CCCCAAGTCAAGAGAGAAATGAGTTATTTTGGCTCTGAAATGCAAAACAGGTTTC 1593  
 QY 461 ArgHisLeuTyrgIuLeuThzSerIleLeuLygGluSerLygTyraIatPserSerGly 480  
 Db 1594 CGCTATTTATACAAATTAATACATCTATTTAAAGAAACAAATTAATACATCGATGCT 1653  
 QY 481 GlyLeuProAlaProSerAspPheLygCysProIleLygIuGluIleAlaIleThzSer 500  
 Db 1654 GGGCGCTGCTCCCAAGGATTTCAAGTCTCTATCAAGAAGAGATGCAATTAATCCAGT 1713  
 QY 501 GlyIuTyrProGluValIleGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520  
 Db 1714 GGTAAATGGAAAGTTCTTGGCCGGCATGATCTAATATCCAAAGTTGATGAAGTCAGAAAG 1773  
 QY 521 LeuValTyrPheGluGluTyrgIuLeuAspSerProLeuGluHisIleLeuTyraIatPser 540  
 Db 1774 CTGGTATATTTTGAAGGACCAAGACATCCCTTTAGAGCATCACCTGTAGCTAGTCACT 1833  
 QY 541 TyrValAsnProGlyGluValIleThzArgLeuThzAspArgIleTyrgIuSerHisSerCys 560  
 Db 1834 TACGTAATCTCTGAGAGAGGTGACAAAGCTGACTGACCTGCTACTACATTTCTTGCTGC 1893  
 QY 561 IleSerGlnHisCysAspPhePheIleSerLygTyrgIuSerAsnGlnLygAsnProHisCys 580  
 Db 1894 ATCAGTACAGACCTGATCTTTTATATAGTAAGTATACAAACAGAAATCCACACACTGT 1953  
 QY 581 ValSerLeuTyrgIuLeuSerSerProGluAspAspProThrCysLygTyrgIuGluPhe 600  
 Db 1954 GTGTCCCTTTCAAGGCTATCAAGTCTCTGAAGATGACCCAACTTGCAAAACAAGGAATTT 2013  
 QY 601 TyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrgIuThzProGluIlePhe 620  
 Db 2014 TGGGCCACCAATTTTGATTCAGCAGGTCTCTCTGCTGACTATACCTCCAGAAATTTTC 2073  
 QY 621 SerPheGluSerThrThzGlyPheThzLeuTyrgIuMetLeuTyrgIuSerProHisAspLeu 640  
 Db 2074 TCTTTAAAGTACTACTGATTAATCATGTATGAGATGCTCTCAAGCCCTCATGATCTTA 2133

QY 641 GlnProGlyLygLygTyrgIuThzProThrValLeuPheIleTyrgIuGly--ProGlnValGlnL 660  
 Db 2134 CAGCTGGAAAGAAATATCTTACTGTCTGCTCATATATGGTGGTCTCTCAGGTGCACT 2193  
 QY 660 euValAsnAsnArgPheLygIleValLygTyrgIuPheArgLeuAsnThzLeuAlaSerLeuG 680  
 Db 2194 TGGTAATTAATCCGTTTAAAGAGTCAAGTATTTCCGTTGAATACCGTACGCTCTCAG 2253  
 QY 680 1TyrgIuValIleValIleAspAsnArgGlySerCysHisArgGlyLeuLygPheGluG 700  
 Db 2254 GTTATGTGTTGTGTGATGATGACAAACAGGGGATCTCTGACAGGGGCTTAAATTTGAAG 2313  
 QY 700 1ValAlaPheLygTyrgIuMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnT 720  
 Db 2314 CGGCTTTAAATTAATTAATGAGTCAAAATGAAATTAAGATCAGATCAGGTGAAGGACTCCAA 2373  
 QY 720 TyrLeuAlaSerArgTyrgIuAspPheIleAspLeuAspArgValGlyIleIleIleIleIle 740  
 Db 2374 ATCTAGCTTCTCGATATGATTTCAATGACTTAATCGTGTGGCATCCAGCGCTGTGCTCT 2433  
 QY 740 TyrgIuGlyTyrgIuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI 760  
 Db 2434 ATGAGAGATACCTCTCTCCGATGAGCATTAATGACAGAGTCAAGATATCTTCAAGGTTGCTA 2493  
 QY 760 LeuAlaGlyAlaProValThrLeuThrIlePheTyrgIuAspThrGlyTyrgIuThzGluArgTyrg 780  
 Db 2494 TTGCTGGGGCCCCAGTCACTGTGTGATCTTCTATGATACAGATACAGAAACCTTATTA 2553  
 QY 780 etGlyHisProAspGlnAsnGluGlnIleTyrgIuTyrgIuGlySerValAlaMetGlnAlaG 800  
 Db 2554 TGGGTCACCTCGACCAAGATGAACAGGCTATTTAGATTCGTGTGGCATGCAAGCAG 2613  
 QY 800 1uLygPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnV 820  
 Db 2614 AAAAGTTCCCTCGAAACCAATCGTTTACGTCTTACATGAGTTTCTCGATGAGAAATG 2673  
 QY 820 aHisPheAlaHisThzSerIleLeuLeuSerPheLeuValArgAlaGlyLygProTyra 840  
 Db 2674 TCATTTTTCACATACCAAGTATATTAATGAGTATTTTAAAGTGGCTGGAAAGCCATATG 2733  
 QY 840 sPLeuGlnIleTyrgIuGlnIleuArgHisSerIleArgValIProGluSerGlyGluHisT 860  
 Db 2734 ATTTACAGATCTATCTCAGAGAGACACACACATTAAGAGTTCCGTAATCCGGAGAACAT 2793  
 QY 860 TyrgIuLeuHisLeuLeuHisTyrgIuGlnIleuAsnLeuGlySerArgIleAlaAlaLeuL 880  
 Db 2794 ATGAACCTGCATCTTTTGGACATACCTCAAGAAACCTTGATCAAGTATGTGCTGCTTA 2853  
 QY 880 ySValIle 882  
 Db 2854 AAGTGATA 2861

RESULT 6  
 US-09-976-674-22  
 ; Sequence 2g, Application US/09976674  
 ; Patent No. US20020115843A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Qi, Steve  
 ; APPLICANT: Akinsanya, Karen  
 ; APPLICANT: Riviere, Pierre  
 ; APPLICANT: Junier, Jean-Louis  
 ; TITLE OR INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
 ; FILE REFERENCE: 70669  
 ; CURRENT APPLICATION NUMBER: US/09/976,674  
 ; PRIOR FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: US 60/240,117  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 22  
 ; LENGTH: 4685  
 ; TYPE: DNA

ORGANISM: Homo sapiens  
US-09-976-674-22

## Alignment Scores:

Pred. No.:	0	Length:	4685
Score:	4385.50	Matches:	834
Percent Similarity:	94.564	Conservative:	0
Best Local Similarity:	94.564	Mismatches:	1
Query Match:	93.314	Indels:	48
DB:	10	Gaps:	1

US-10-070-464-1 (1-882) x US-09-976-674-22 (1-4685)

```

QY      1 MetAlaAlaIaMeCgluThrGluGluGluValGluIlePheGluThrAlaAspCys 20
DB      214 ATGGCAGCAGCATGGAACAGACAGCGGGTGTGAGATATTGAACTCGGACTGT 273
QY      21 GluGluAsnIleGluSerGluAspArgProLysLeuGluProPheTyrValGluArgTyr 40
DB      274 GAGGAGAAATATGCAATCACAGGATCGGCTAAATTGAGACCTTTTATGTGAGCGGTAT 333
QY      41 SerTPSerGluLeuLysLeuLeuAlaAspThrArgLeuTyrHisGlyTyrMetMet 60
DB      334 TCTTGAGTCAAGCTTAAAGAGCTTGCCGATACCAAGAAATATCATGGCTATCATGATG 393
QY      61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
DB      394 GCTAAGGACCAACATGATTTGATGTTGGAAGAGAAATGATCCAGATGGACCTCATTTCA 453
QY      81 AspaArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB      454 GACAGATCTATTAATCACTTCATGCTGCTGAGAAACAGAGAAATACACTGTTTATTTCT 513
QY      101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
DB      514 GAAATTCCTCAAACTATCAATAGACAGACAGACTTTAATCTCTCTTGGAAGCCCTCTTTG 573
QY      121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
DB      574 GATCTTTTCAAGCAACACTGGACCTATGGAATGTATTTCTCGAAGAAAGAACTATTAGA 633
QY      141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB      634 GAAAGAAAACGATTTGGAACAGTCGGAATGCTCTTACGATTATTCACCAAGAAAGTGA 693
QY      161 ThrPheLeuPheGluAlaGlySerGlyIleTyrHisValLysAspGlyValProGlnGly 180
DB      694 ACATTTCTGTTCCAAACCCGGTAGTGAAATTTATCACGTTAAAGATGAGAGGCCACAGGA 753
QY      181 PheThrGlnGluProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB      754 TTACGCAACCAACTTTAAGGCCCAATCTAGTGAAACTAGTTGTCCTCCAACTACGGATG 813
QY      201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
DB      814 GATCCAAATTAATGACCTGCTGATCCAGACTGGAATGCTTTTATCATGTGCAACCATATT 873
QY      221 TrpIleSerAsnIleValThrArgGluGluArgGlyLeuThrTyrValHisAsnGluLeu 240
DB      874 TGGATATCTAATCATGCTAATCCAGAGAGAGAAAGAAAGAACTACTTATAGTCACATAGCT 933
QY      241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
DB      934 GCCAACAATGGAAGAGATGCCAGATCAGCTGAGTGGCTACCTTTGTTCTCCAAAGAGA 993
QY      261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
DB      994 TTTGATAGATATTCTGGCTATGGGTATGGGTCCAAAGCTGAAACAACCTCCAGTGTGAT 1053
QY      281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB      1054 AAAATTTCTAGAAATCTATATGAGAAATATGATGATCGAGGTGGAATTTATCATGTT 1113

```

```

QY      301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB      1114 ACATCCCTATGTTGGAAACAGAGAGGCGAGTTTATTCCTGTTATCTTAAACAGGTACA 1173
QY      321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluValArgIle 340
DB      1174 GCAAATTCCTAAAGTCACTTTTAAAGATGTACAGAAATTAATGATGATGCTGAAGAGAGATC 1233
QY      341 IleAspValIleAspLysGluLeuLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
DB      1234 ATGATGTCATATGATTAAGAAAGAACTAATCACTTTTGAGATTTCTATTGAAAGAGTTGAA 1293
QY      361 TyrIleAlaArgAlaGlyTyrTrpProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
DB      1294 TATATTCGACAGAGCTGATGATGACTCTCAGAGGAAATATGCTTGCTCATCTACTAGAT 1353
QY      381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB      1354 CGCTCCAGACTCGCTCAGATAGTGTGATCTCACTGAATTTATTATCCAGTAGAA 1413
QY      401 AspaAspValMetGluArgGluAlaArgLeuIleGluSerValProAspSerValThrProLeu 420
DB      1414 GATGATGTTATGGAAGGACAGACTCATTAAGTACGCTGATTTCTGTGACGCCACTCA 1473
QY      421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
DB      1474 ATTATCTATGAAGAAACAAACAGACATCTGGATTAATATCCATGACATCTTTTATGTTT 1533
QY      441 ProGlnSerHisGlyGluGluIleGluPheIlePheAlaSerGlyCysLysThrGlyPhe 460
DB      1534 CCCCAAGTCAAGAAAGAAATTAAGTTTATTTTGTGCTCGAATGCAAAACAGATTTTC 1593
QY      461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerGly 480
DB      1594 GCTATTTATCAAAATTAATCACTATTAAAGGAAACCAATATAAAGCATCCAGTGT 1653
QY      481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
DB      1654 GGGCGCTGCTGCCAAGGATTTTCAAGTGTCTTATCAAAAGAGAGATACCAATTTCCACT 1713
QY      501 GlyIleTrpGluValLeuGlyValArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB      1714 GGTGAATGGAAGATTTCTTGCGCGGATGATCTAATATCAAAATGATGAAGTCAAGAGG 1773
QY      521 LeuValTyrPheGluGluThrLysAspSerProLeuGlnHisIleLeuTyrValIleSer 540
DB      1774 CTGGTATATTTTGAAGGCAACAAAGACTCCCTTTAAGACATCACCTGTACGTAGTCACT 1833
QY      541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
DB      1834 TACGTAAATCTCGAGAGAGTACAAAGCTGACTGACCGTGGCTACTCATTCTTGCTGC 1893
QY      561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
DB      1894 ATCAGTCAGCACTGGAATCTTTATATAGTAAATAGTAACTCAAGAAATCCACACTGT 1953
QY      581 ValSerLeuTyrLysLeuSerSerProGluAspAspProTrpCysLysThrLysGluPhe 600
DB      1954 GTGTCCCTTTCAAGACTATCAAGTCCGAAAGATGACCAACTTGCACAAAGAAATTT 2013
QY      601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
DB      2014 TGGGCAACATTTTGATTTGATTCGT----- 2036
QY      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
DB      2036 ----- 2036
QY      641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGlnValGlnLeu 660
DB      2037 -----CCTCAGATCAGATG 2051
QY      661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

```

Accession	Gene	Protein	Length	Score
Db	2052	GTGATTAATACGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTAGCTTTAGGT	2111	4385.00
QY	681	TYRVALVALVALVALLAASPANARGISERYSHISARGILEULYPHEGLUGLY	700	831
Db	2112	TATGTGTTGTAGTGATAGACAAACAGGGATCTCTGCACCGAGGCTTAAATTTGAAGC	2171	
QY	701	ALAIPHELYTYRILEYMEGLYGLNILEGUILAEPAPGALVALGLUGLYLEUGINTYR	720	
Db	2172	GCCTTTAAATATATAAATGGGTCAATATAGAAATGTACAGTACAGTGAAGGACTCCAAATAT	2231	
QY	721	LEUALASERARGTYRASPHEILEASPLEUASPAVALGLYLEHISGLYTRSERITYR	740	
Db	2232	CTAGCTTCTCGATATGATATTTCTTGACTTAGATCGTGGGCATACAGCGCTGCTCTAT	2291	
QY	741	GLYGLYTYRLEUSERLEUWEALAEUWEGINARGSERAPLIPHEARGVALAILE	760	
Db	2292	GGAGGATACCTCTCCCTGATGGCAATTAATGCAGAGTCAGANATCTTAGGGTTGCTATT	2351	
QY	761	ALAGLYALAPROVALTHLEUTRPILEPHELYRASPTRNGLYTYRTHGLUARGTYRMEC	780	
Db	2352	GCTGGGGCCCCAGTCACTTGTGATCTTCTATGTATACAGGTACACGGAAGCTTATATG	2411	
QY	781	GLYHISPROASPGALASNGLUGLNGLYTYRTRYLEUGISERYVALAILEGLALAGLU	800	
Db	2412	GGTCACTCTGACCAAGATGACAGGGCTATTTACTTAGATCTGTGGCCATGACGACAGAA	2471	
QY	801	LYSPHEPROSEGLUPROASNARGLEULEULEUHLISGLYPHELEUASPGILUASVAL	820	
Db	2472	AAGTTCCCTCTGAAACCAATGTTTACGTGCTTCAATGGTTTCTTGAGAGAAATGTC	2531	
QY	821	HISPEALAHISTHISERILEULEUSERPHELEUVALAAGLALGLYLYSPOTYRASP	840	
Db	2532	CATTTTGCACATACACAGATATATTTACTGTAGTTTTTATGTAGAGGGCTGGAACCCATATGAT	2591	
QY	841	LEUGNILEYRPRGGLNGLUARGHISERILEARGVALPROGLUSERGLYGLNILEYR	860	
Db	2592	TTACAGATCTATCTCAGAGAGACACAGCATMAGATTTCCGTAATCGGAGAACATTAT	2651	
QY	861	GLUEUHLISLEUHLISTRYLEUGLNGLUASNLEUGISERYSERARGILEALALEULYS	880	
Db	2652	GAACTGCATCTTTTGCACCTTCAABAAAACTTGATCATCGATTTGCTGCTTAAAA	2711	
QY	881	VALIIE 882		
Db	2712	GTGATA 2717		
RESULT 7				
US-09-976-674-20				
; Sequence 20, Application US/09976674				
; Patent No. US20020115843A1				
; GENERAL INFORMATION:				
; APPLICANT: Qi, Steve				
; APPLICANT: Akinsanya, Karen				
; APPLICANT: Riviere, Pierre				
; APPLICANT: Junliem, Jean-Louis				
; FILE REFERENCE: 70669				
; CURRENT APPLICATION NUMBER: US/09/976,674				
; PRIOR FILING DATE: 2001-10-12				
; PRIOR APPLICATION NUMBER: US 60/240,117				
; NUMBER OF SEQ ID NOS: 61				
; SOFTWARE: PatentIn version 3.1				
; SEQ ID NO 20				
; LENGTH: 4676				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
US-09-976-674-20				
Alignment Scores:				
Pred. No.: 0				
Score: 4385.00				
Length: 4676				
Matches: 831				

Percent Similarity:	94.22%	Conservative:	0
Best Local Similarity:	94.22%	Mismatches:	1
Query Match:	93.30%	Indels:	51
DB:	10	Gaps:	1
US-10-070-464-1 (1-882) x US-09-976-674-20 (1-4676)			
QY 1 MetAlaAlaIaMeCgIuThrGluIuLeuGlyValGluIlePheGluThraIaAspCys 20			
DB 214 ATGGGAGCGAGCAATGAAACAAACAACAGCTGGGCTTGAGATATTGAAATCGCGAATGT 273			
QY 21 GluGluAsnIleGluSerGlnAspArgProGlySLeuGluProPheTyrValGluArgTyr 40			
DB 274 GAGGAGAAATTAATTAATCAAGATCGGCTAAATTGAGCCTTTTAATGTTGACCGGTAT 333			
QY 41 SerTPSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60			
DB 334 TCCTGAGTCAGACTTAAAGCTGCTTCCCGATACAGAAATATATATGAGCTCATGATG 393			
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80			
DB 394 GCTAAGGCGACCAATGATTTCTGTTGAAAGGAATGATTCAGATGAGCTCATTTCA 453			
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyLysAsnArgLysAsnThrLeuPheTyrSer 100			
DB 454 GACAGAAATCTATTACCTTGCCATGCTGGTGAGAACAGAAATACACTGTTTATCT 513			
QY 101 GluIleProLysThrIleAsnArgAlaAlaValIleMetLeuSerTyrLysProLeuLeu 120			
DB 514 GAAATTCCTCCAAACTATCATATGACGACGACGACTTATCTCTTGAGAGCCTCTTTTG 573			
QY 121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140			
DB 574 GATCTTTTTCAGCAACAACCTGACTATGGAATGTATATTCGAGAAAGAACTATTAA 633			
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160			
DB 634 GAAAGAAAAACGATTTGAAACAGCTCGGAATTCCTTTCAGATTATCCAAAGAGAGTGA 693			
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180			
DB 694 ACATTTCTGTTTCAAGCCGGTATGTGAATTTATCCGTTAAAGATGAGAGGCGCACAAAG 753			
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200			
DB 754 TTACGCAACAACCTTTAAAGCCCAATCTAGTGAACCTAGTTGCCAACATACCGAATG 813			
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220			
DB 814 GATCCAAATTAATGCCCTGCTGATCCAACTGATTTGCTTTATACATAGCAACGATATT 873			
QY 221 TrpIleSerAsnIleValThrArgGluGluValArgLysLeuThrTyrValHisAsnGluLeu 240			
DB 874 TGGATATTTACATCTGTAAACCAAGAAAGAAAGAAACCTCATTTATGTGCAACATGAGTA 933			
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260			
DB 934 GCCAAACATGAGAAAGATGCCGAGATCACTGGAGTCGCTGTTGTTCCAGAAAGAA 993			
QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280			
DB 994 TTGATATATATTTCTGCTATTTGGTGCTGCCAAAGCTGAACAACATCCCACTGGTGT 105			
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300			
DB 1054 AAAATTTCTTAAGATTTCTATGAAAGAAATGTGAATCTGAGGTGAATTAATTAATGTT 111			
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320			
DB 1114 ACATTCCTCATGTTGTAAGAAACAAGAGGCGCAATTCATTCCTGTTATCTTAAACAGTACA 1177			
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyValGlyLe 340			

Db	1174	GCMAATCTCTAAGTACCTTTTAAGATGTCAGAAATAATGATTGATGCTGAAAGGAAGATC	1235
Qy	341	IIeaSPValIIeaSPLySGluLeuIleGlnProPheGluIleLeuPheGluIValGlu	360
Db	1234	ATGATGTCATGATATAGAACTTAATCACTTTTGAGATTCTATTGGAAGAGTTGA	1292
Qy	361	TyrIIeaIaArgAlaGlyTrpThrProGluGlyIleValAlaTrpSerIleLeuLeuAsp	380
Db	1294	TATATTGCCAGGCTGGATGAGCTCCTGAGGAAATATGTTGGTCCATCCTACTTAAT	1353
Qy	381	ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu	400
Db	1354	CGCTCCAGACTCGGCTACAGATAGGTTGATCTCACCTGAATTATTTATCCAGATAGA	1413
Qy	401	AspAspValMetGlnArgGlnArgLeuIleGluSerValProAspSerValThrProLeu	420
Db	1414	GATGATGTTATAGAAAGCGAGACTCATGTAGTCAAGGCCGATTCGTGTACGCCACTA	1473
Qy	421	IleIleIleArgGluGlnThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe	440
Db	1474	ATTATCTATGAAGAAACAACAGACATCTGGATAATATCCATGACATCTTTCATGTTT	1533
Qy	441	ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIleThrGlyPhe	460
Db	1534	CCCCAAAGTACGAAAGAGAAATTGATTTATTTTGGCTCTGAATGCAAAACAGTTTC	1593
Qy	461	ArgHisIleLeuIleThrIleIleThrSerIleLeuIleGlySerIleIleValArgSerSerGly	480
Db	1594	CGTCATTTATACAAAATTACATCTATTTTAAAGAAAGCAAAATATTAACGATCCAGTGT	1653
Qy	481	GlyLeuProAlaProSerAspPheIleCysProIleIleGluGluIleAlaIleThrSer	500
Db	1654	GGCGTCCGCTGCCAAGATGATTCAAGTGCTCTATCAAAAGAGAGATAGCAATTACAGT	1713
Qy	501	GlyGluTrpGluValLeuGluValArgHisGlySerAsnIleGlnValArgGluValArgArg	520
Db	1714	GGTGATAGGGAAGTCTTGCGCGGATGATTAATATTCACAAGTGAAGATCGAAGTCAAGG	1773
Qy	521	LeuValIleIlePheGluGlyThrIleAspSerProLeuGluHisIleIleLeuIleValIleSer	540
Db	1774	CTGGTATATTTTGAAGGCAACCAAGACTCCCTTTAGAGATCACCTGTAGCTAGTCACT	1833
Qy	541	TyrValAsnProGluValIleThrArgLeuThrAspArgGlyIleSerHisSerCysCys	560
Db	1834	TACGTAAATCCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCCATCTTGCTGC	1893
Qy	561	IleSerGlnHisCysAspPhePheIleSerIleIleIleIleIleIleIleIleIleIleIleIle	580
Db	1894	ATCAGTCAAGCACTGAGACTCTTTATTAAGTAAAGTATAGTAAACCAAGAAATCCACACTGT	1953
Qy	581	ValSerLeuIleIleLeuSerSerProGluAspAspProIleCysIleThrIleGluIlePhe	600
Db	1954	GTGTCCCTTTACAAAGCTATCAAGTCCCTGAAGATGACCAACCTTGCAAAACAAAGGAATTT	2013
Qy	601	TrpAlaTrpIleLeuAspSerIleAlaIleProLeuProAspIleIleIleIleIleIleIlePhe	620
Db	2014	TGGGCAACCATTTTGATTCAGACAGTCTCTCTCTGACTAATACCTCTCCGAAATTTTC	2073
Qy	621	SerPheGlnSerThrThrGlyPheThrIleLeuIleGlyMetLeuIleIleIleIleIleIleIleIle	640
Db	2074	TCTTTTGAAGAGTACTAGTGAATTTCATTTGATAGTAAAGCTCTACAAACCTCATGATCTTA	2133
Qy	641	GlnProGlyIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle	660
Db	2134	CAGCCTGAAAGAAATATCTCATCTGCTGTTCAATATGTGGTGTGC-----	2180
Qy	661	ValAsnAsnArgPheIleGlyValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle	680
Db	2180	-----	2180
Qy	681	TyrValValValValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle	700
Db	2180	-----	2180

QY	701	AlaphehlyRfYrLvSmeGlyGlnlleguileapapGlnValGluGlyLeuGlnTyr	720
Db	2191	-----GGTCAATAGAAAATTGACGATCGAGTGGAGGACTCCAAATAT	2222
QY	721	LeuAlaSerArgTyrAspPheIleAspLeuSpArgValGlyIleHISGLYTrpSerTyr	740
Db	2223	CTAGGCTTCGATATGATTTCTTCATTTGACTTACGTTCGTGGGCATCCAGGCTGGTCTTAT	2288
QY	741	GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle	760
Db	2283	GGAGGATACCTCTCCCTGATGGCATTTATGACAGAGTCAGATATCTTCAGGCTTCTATT	2342
QY	761	AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet	780
Db	2343	GCTGGGGGCCCAAGTACCTCTGTGGATCTCTATGATACAGGATACCGAAAGTTATATG	2402
QY	781	GlyHisProAspGlnAsnGlyGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu	800
Db	2403	GCTACCCCTGACACGAAATGAAACAGGGCTATTACTTATGAGATCTGTGGCCATGCACAGCAA	2466
QY	801	LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal	820
Db	2463	AAGTCCCCCTCGAACCAATCGTTACTGCTCTTACATGGTTTCTCGATGAGAAATGTC	2522
QY	820	HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLysProTyrAsp	840
Db	2523	CATTGTGCATACACAGATATATTACTGTGTTTTAGAGAGCGCTGGAAACCAATATCAT	2582
QY	841	LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlnSerGlyGlnHisTyr	860
Db	2583	TTACGATCTTATCTCTCAGAGAAAGACACGATTAAGAGTTCTCGAATCGGGAGAACATTAT	2642
QY	861	GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys	880
Db	2643	GAACTGCATCTTTGGCACTTACCTTACAGAAAACCTTGATCAGCATTCCTGCTTAATAA	2702
QY	881	ValIleIle882	
Db	2703	GTGATA2708	
RESULT 8			
US-09-976-674-8			
; Sequence 8, Application US/09976674			
; Patent No. US20020115843A1			
; GENERAL INFORMATION:			
; APPLICANT: Qi, Steve			
; APPLICANT: Akinsanya, Karen			
; APPLICANT: Riviere, Pierre			
; APPLICANT: Julien, Jean-Louis			
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV			
; FILE REFERENCE: 70669			
; CURRENT APPLICATION NUMBER: US/09/976,674			
; CURRENT FILING DATE: 2001-10-12			
; PRIOR APPLICATION NUMBER: US 60/240,117			
; PRIOR FILING DATE: 2000-10-12			
; NUMBER OF SEQ ID NOS: 61			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 8			
; LENGTH: 4523			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-976-674-8			
Alignment Scores:			
Pred. No.: 0			
Score: 4092.50			
Percent Similarity: 88.44%			
Best Local Similarity: 88.44%			
Query Match: 87.07%			
DB: 10			
Length: 4523			
Matches: 780			
Conservative: 0			
Mismatch: 1			
Indels: 102			
Gaps: 1			
US-10-070-464-1 (1-882) X US-09-976-674-8 (1-4523)			

QY 1 MetAlaAlaMetGluThrGluGluLeuGlyValGlu1 ephGluThrAlaAspCys 20  
 Db 214 ATGGCAGCGCAATGGAAACAGACAGCTGGGTGGATATTTGAACTGGCGACTGT 273  
 QY 21 GluGluAsn1LeuSerGluAspArgProLysLeuGluProPheTyrValGluArgTyr 40  
 Db 274 GAGGAAATATTGAAATCACAGAGATCGGCTAAATTGAGACCTTTTATGTGTGACGGTAT 333  
 QY 41 SerTPSerGluLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60  
 Db 334 TCCGTGAGTCAGCTTAAAGCTGTGGCTTGCAGATACAGAAATATATGCTTACATGAGT 393  
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAspAspProAspGlyProHisSer 80  
 Db 394 GCTAAGGCAACACATGATTTTCATGTTTGTGMAAGAAAGATCCAGATGGACTCTTCA 453  
 QY 81 AspArg1LeTyrTyrLeuAlaMetSerGlyGluAsnArgLysAsnThrLeuPheTyrSer 100  
 Db 454 GACAGAACTATTAACCTTCCCATGTCTGGTGAGAAACAGAAATAACACTGTTTATCT 513  
 QY 101 Glu1LeuProLysThr1LeaAspArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120  
 Db 514 GAAATTTCCCAAACTATCAATAGAGAGAGAGCTTAAATGCTCTTGGAAAGCTCTTTTG 573  
 QY 121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140  
 Db 574 GATCTTTTTCAGGCAACACTGGACCTATGGAAATGATATCTCGAAGAAAGAACTATTAA 633  
 QY 141 GluArgLysArg1LeGlyThrValGly1Lea1LeaSerTyrAspTyrHisGlnGlySerGly 160  
 Db 634 GAAAGAAACGCAATGGAAACAGTCGGAATTCCTTACGATTTATACCAAGAAAGATGGA 693  
 QY 161 ThrPheLeuPheGluAlaGlySerGly1LeTyrHisValLysAspGlyGlyProGlnGly 180  
 Db 694 ACATTTCTGTTTCAAGCCGTAAGTGGAAATTTATCACGTAAGAGTGAAGGCCCAAGGA 753  
 QY 181 PheThrGlnGluProLeuArgProAsnLeuValGluThrSerCysProAsn1LeaArgMet 200  
 Db 754 TTTACGCAACCACTTTAAGCCCAATCTAGTGGAAACTAGTTGTCCTCAACTACGAGAT 813  
 QY 201 AspProLysLeuCysProAlaAspProAspTrp1Lea1Phe1LeHisSerAsnAsp1Le 220  
 Db 814 GATCCAAATTTATGCCCTGCTGATCCAGACTGGATTCCTTTATACATGCAACGATATT 873  
 QY 221 Trp1LeSerAsn1LeValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240  
 Db 874 TGGATATCTAACATGCTAAACAGAGAAAGAGAAAGAACTCATATATGTGACAAATGACT 933  
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAla1ThrPheValLeuGlnGluGlu 260  
 Db 934 GCCAACATGGAAAGAAAGATGCCAGATCGAGTGGAGTGGTACTTTGTTCTCCAAAGAA 993  
 QY 261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysValGluThrThrProSerGlyGly 280  
 Db 994 TTTGTAGATATTCTGGCTATGTGGTGTCTCCAAAGCTGAAACAACTCCCACTGGTGT 1053  
 QY 281 Lys1LeuLeuArg1LeuLeuTyrGluGluAsnAspGluSerGluValGlu1Le1LeHisVal 300  
 Db 1054 AAAATTTTAGAATTTCTATATAGAGAAATGATGAAATCGAGGTGAAATTTATCATGTT 1113  
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyTyr 320  
 Db 1114 ACATCCCTCATGTTGGAAACAGAGAGGAGATTCATTCCTTTATCCCTTAAACAGGTACA 1173  
 QY 321 AlaAsnProLysVal1ThrPheLysMetSerGlu1LeMet1LeAspAlaGluGlyArg1Le 340  
 Db 1174 GCAAAATCTTAAGTCACTTTTAAAGATGTCAGAAATTAATGATGATGCTGMAAGAAAGATC 1233  
 QY 341 IleAspVal1LeAspLysGluLeuLeu1LeGlnProPheGlu1LeLeuPheGluGlyValGlu 360  
 Db 1234 ATAAGTGTCAATGATGAAGAACTAATTCACCTTTTGAATTCATTTTGAAGAAATTGA 1293

QY 361 Tyr1Lea1LeArgAlaGlyTrpThrProGlnGlyLysTyrVala1TrpSer1LeLeuLeuAsp 380  
 Db 1294 TATATTGCCAGAGCTGATGAGACTCTGAGGGGAAATATGCTTGGTCCATCTCACTAGAT 1353  
 QY 381 ArgSerGlnThrArgLeuGln1LeVal1Leu1LeSerProGluLeuPhe1LeProValGlu 400  
 Db 1354 CCTCCCAACACTCGCTTACAGATAGTGTGATCTCACTGAAATTAATTTATCCAGATGA 1413  
 QY 401 AspAspValMetGluArg1LeArgLeu1LeGluSerValProAspSerValThrProLeu 420  
 Db 1414 GATGATGTTATGAAAGGAGAGACTCATTTAGTCACTGCTGATTTCTGAGCGCACTA 1473  
 QY 421 Ile1LeTyrGluGluThrThrAsp1LeTrp1LeAsn1LeHisAsp1LePheHisValPhe 440  
 Db 1474 ATTATCTAAGAAAGAAACAAACAGACATCTGATTAATATCAATACACTCTTTCATGTTT 1533  
 QY 441 ProGlnSerHisGluGluGlu1LeGluPhe1LePhe1LeSerGluCysLysThrGlyPhe 460  
 Db 1534 CCCCAAGTCAGAAAGAGAAATTGAGTTATTTTGGCTTGAAATGCAAAACAGGTTTC 1593  
 QY 461 ArgHisLeuTyrLys1LeThrSer1LeLeuLysGluSerLysTyrLysArgSerSerGly 480  
 Db 1594 CGCATTTATACAAATTAATCATTTTAAAGGAAAGAAATATTAACGATCCAGTGT 1653  
 QY 481 GlyLeuProAlaProSerAspPheLysCysPro1LeLysGluGlu1LeAla1LeThrSer 500  
 Db 1654 GGGCTCCCTCGCTCCAAAGTATTTCAAGTCTCTATTAAGAGAGATAGCAATTAACAGT 1713  
 QY 501 GlyGluTrpGluValLeuGlyValArgHisGlySerAsn1LeGlnValAspGluValArgArg 520  
 Db 1714 GGTGAATGGGAAGTTCTTGCGCGCATGGATCTTAATATCCAAATGATGAAGTCAGAA 1773  
 QY 521 LeuVal1TyrPheGluGlu1TyrThrLysAspSerProLeuGluHisGlyLeuTyrVal1ValSer 540  
 Db 1774 CTGGTATATTTGAAAGGCAACAAAGCTCCCTTTAGACATCACTGTAACGATGACT 1833  
 QY 541 TyrValAsnProGlyGlyVal1ThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560  
 Db 1834 TAGGTAAATCCCGAGAGGTGACAAAGCTGATGACCGTGGCTACTACATCTTCTCTC 1893  
 QY 561 HisSerGlnHisCysAspPhePhe1LeSerLysTyrSerAsnGlnLysAsnProHisCys 580  
 Db 1894 ATCAGTCAAGCACTGAGACTTTCTTATAGTAAGTATAGTAACAGAAAGATCCACACTGT 1953  
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600  
 Db 1954 GTGTCCCTTTACAACTATCAAGTCCGAAAGATGACCCCACTTGCAAAACAAAGGAATT 2013  
 QY 601 TrpAlaThr1LeLeuAspSerAlaGlyProLeuProAspTyrThrProGln1LePhe 620  
 Db 2014 TGGGCCACCATTTTGGATTCAGCAGGCTCTCTTCTGACTATATCTCTCCAGAAATTTTC 2073  
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640  
 Db 2074 TCTTTTGAAGTACTACTGATTTTACATGTATGGATGCTCTCAACAGCTCATGATCTTA 2133  
 QY 641 GlnProGlyLysLysTyrProThrValLeuPhe1LeTyrGlyGlyProGlnValGlnLeu 660  
 Db 2134 CAGCTGGAAGAAATATCTCACTGTGCTGATATAGTGGT----- 2178  
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680  
 Db 2178 ----- 2178  
 QY 681 TyrValValVal1LeAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700  
 Db 2178 ----- 2178  
 QY 701 AlaPheLysTyrLysMetGlyGln1LeGlu1LeAspAspGlnValGluGlyLeuGlnTyr 720  
 Db 2178 ----- 2178  
 QY 721 LeuAlaSerArgTyrAspPhe1LeAspLeuAspArgValGly1LeHisGlyTyrPheSerTyr 740



Db 2178 ----- 2178  
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760  
Db 2179 ----- -CGGT- GCTATT 2189  
QY 761 AlaGValAlaProValThrLeuTrrPilePheTyrAspThrGlyTyrThrGluArgTyrMet 780  
Db 2190 GCTGGGGCCCCAGTCACTGTGTGATCTTCTATGATACAGGATACAGGAAAGCTTATATG 2249  
QY 781 GlyHisProAspGlnaEngIuGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800  
Db 2250 GGTCAACCTTCAGCAGATGAACAGGCTATTACTAGGATCTGTGGCATTGCAAGCAGAA 2309  
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820  
Db 2310 AAGTTCCTCCCTCAACCAATCGTTTACTCTCTTACATGTGTTCCCGGATGAGAAATGTC 2369  
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840  
Db 2370 CATTTTGACATATCCAGTATATTAATCTAGATTTTATTAGAGGCTGGAAAGCCATATGAT 2429  
QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860  
Db 2430 TTACAGATCTATCTCTGAGAGAGACACAGCATTAAGATTCCTGAATCGGAGAACTAT 2489  
QY 861 GluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880  
Db 2490 GAACCTGCATCTTTTGCACTACCTTCAGAAACCTTGATGACGATATTCCTGCTCTMAAA 2549  
QY 881 ValIle 882  
Db 2550 GTGATA 2555

RESULT 9  
US-09-976-674-14 : Sequence 14, Application US/0976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Aktinsanya, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 14  
LENGTH: 4309  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-976-674-14

Alignment Scores:  
Pred. No.: 0 Length: 4309  
Score: 3661.50 Matches: 708  
Percent Similarity: 80.09% Conservative: 0  
Best Local Similarity: 80.09% Mismatches: 1  
Query Match: 77.90% Indels: 176  
DB: 10 Gaps: 1

US-10-070-464-1 (1-882) x US-09-976-674-14 (1-4309)

QY 1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20  
Db 214 ATGGCAGCAGCATGGAAGAAAGACAGCTGGCTTGAGATATTTGAAACTGGCGACTGT 273  
QY 21 GluGluAsnIleGluSerGlnaAspArgProLysLeuGluProPheTyrValGluArgTyr 40

0  
Db 274 GAGGAGAAATTTGAATCACAGGATCGGCTTAATTGGAGCTTTTATGTGACGGTAT 333  
QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60  
Db 334 TCCTGGAGTCAGCTTAAAAAGCTGCTGCCGATACCAAGAAAATATCATGGCTACATGATG 393  
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80  
Db 394 GCTAAAGCACACATATATTTCAATGTTGTGMAAGAAATGATCAATGAGACCTCATTTCA 453  
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgLysAsnThrLeuPheTyrSer 100  
Db 454 GACAGAAATCTATTAATCTTGCATGTCTGGTGAACAACAGAAAATACATCTTTTATCT 513  
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrrLysProLeuLeu 120  
Db 514 GAATTCCTCCAAACATATCAATAGACAGCAGCTTATATGCTCTTGAAGCCCTCTTTTG 573  
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140  
Db 574 GATCTTTTTCAGGCAACACTGACCTATGAAATGATATTCGAGAAAGAACTATTAAAGA 633  
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160  
Db 634 GAAGAAAGAAAGCATTTGGAAACAGTCGGAATTCCTTTACATTAATCACCAAGAAAGTGA 693  
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180  
Db 694 ACATTTCTGTTTCAACCGGTAGTGAATTTATCAGTAAAGATGAGAGGCCCAAGAGA 753  
QY 181 PheThrGlnGluProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200  
Db 754 TTACGCCAACACCTTTAAAGCCCAATCTAGTGAATTAATGTTGCTCCCAACATACGAGAT 813  
QY 201 AspProLysLeuCysProAlaAspProAspTrrPileAlaPheIleHisSerAsnAspIle 220  
Db 818 GATCCAAATTAATGCCCCCTGCTGATCCAGACTGATGATTCCTTTTATCATAGCAACGATATT 873  
QY 221 TrrPileSerAsnIleValThrArgGluGluArgGluLeuThrTyrValHisAsnGluLeu 240  
Db 874 TGGATATCTTAACATCGTAAACAGAGAAGAAAGAGACTCACTATGTCACATAGACTA 933  
QY 241 AlaAsnMetGluGluAsnAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260  
Db 934 GCCAACATGGAAAGAAATGCGCAGATCAGCTGGAGTGTCTTATCTTCCAAAGAGAA 993  
QY 261 PheAspArgTyrSerGlyTyrTrrPTrCysProLysAlaGluThrThrProSerGlyGly 280  
Db 994 TTTGATAGATATTTCTGGCTATTTGGTGTGTCCTCAAAAGCTGAAACACTCCCACTGGTGT 1053  
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300  
Db 1054 AAAATCTTAAGATTTCTATATGAAAGAAATGATGATCTGAGGTGAAATTTATTCATGTT 1113  
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320  
Db 1114 ACATCCCTATATGTGAAACCAAGAGGCGAGATTCCTCGTTATCTTAAACAGGTATCA 1173  
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgGly 340  
Db 1174 GCAAACTCCAAAGTCACTTTTAAGATGTCAGAAATTAATGATGATCTGAAAGAAATATC- 1232  
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360  
Db 1232 ----- 1232  
QY 361 TyrIleAlaArgAlaGlyTrrPThrProGluGlyLysTyrAlaTrrPserIleLeuLeuAsp 380  
Db 1232 ----- 1232  
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400



```

Db      1232 ----- 1232
QY      401 AspAspValMetGluArgGlnArgLeuLeuGluSerValProAspSerValThrProLeu
Db      1232 ----- 1232
QY      421 IleIleTyrGluGluThrThrAspIleTyrPheAsnIleHisAspIlePheHisValPhe
Db      1232 ----- 1232
QY      441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIleThrGlyPhe
Db      1232 ----- 1232
QY      461 ArgHisLeuTyrIleThrSerIleLeuLeuGluSerIleTyrIleAspGlySerSerGly
Db      1232 ----- 1232
QY      481 GlyLeuProAlaProSerAspPheIleCysProIleLeuGluGluIleAlaIleThrSer
Db      1232 ----- 1232
QY      501 GlyGluTyrPheGluValLeuGluIleArgHisGlySerAsnIleGlnValAspGluValArg
Db      1233 ----- 1253
QY      521 LeuValTyrPheGluGluIleThrIleAspSerProLeuGluHisIleLeuTyrValValSer
Db      1254 CTGGATATATTTGAAAGCACAAGAACTCCCTTTAGAGCATACCTGATACGTAAGTCAGT 1213
QY      541 TyrValAsnProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys
Db      1314 TACGTAATCTCGAGAGAGGAGCAAGGCTGACGACCGTGGCTACCTACCTTCTGCTGC 1373
QY      561 IleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnIleAsnProHisCys
Db      1374 ATCAGTCACACACTGACTCTTTATTAAGTAATATAGTAACAGAAATCCACACTGT 1433
QY      581 ValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysValSerThrIleGluPhe
Db      1434 GTGTCCCTTTACAGAGTATCAAGTCTGTAAGATGACCCCACTTGCAAAACAAAGAAATTT 1493
QY      601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe
Db      1494 TGGGCACCATTTTGGATTGACAGAGGCTCTTCGACATATACCTCCAGAAATTTTC 1553
QY      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeu
Db      1554 TCTTTGAAAGTACTACGATTTTACATTGTATGGAGTGTCTTACCAAGCCTCATGATCTA 1613
QY      641 GlnProGluIleLeuSerTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGln
Db      1614 CAGCCTGGAAGAAATATCTCTACTGTGCTGTCATATATAGTGCTCTCCACAGTCACT 1673
QY      660 euValAsnAsnArgPheIleGlyValIleTyrPheArgLeuAsnThrLeuAlaSerLeuG
Db      1674 TGGTGAATATATGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTCTAG 1733
QY      680 IlyTyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLeuPheGluG
Db      1734 GTTATGTGTTGTAGATAGACCAACAGGGGATCCCTGTACCGAGGCTTTAAATTTGAAG 1793
QY      700 IlyAlaPheLeuTyrIleSerMetGlyGlnIleGlnIleAspAspGlnValIleGluIleGlnT
Db      1794 GGGCCTTTAAATATATAATGAGTCAATATGAAATATACATCAAGGGAAGGACTCCAAAT 1853
QY      720 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheT
Db      1854 ATCTAGACTTCTCGATATGATTTTCACTTGAATGATGTGTGGGCAATCCACGGCTGTCTCT 1913
QY      740 TyrGlyGlyTyrIleSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI
Db      1914 ATGGAGGATACCTCTCCCTGATGGCATTAATATGACAGGTCAGATATCTTCAGAGTTGGCTA 1973

```

```

QY      760 leAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrM
Db      1974 TTCTGGGGCCCAAGTACACTCTGTGAATCTTCTATGATATACAGATACAGGAAGCTTATA 2033
QY      780 etGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaG
Db      2034 TGGGTACACCTGACCAAGATAGAACAGGGCATTAATAGGATCTGTGGCCATCAGAACAG 2093
QY      800 IulysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnV
Db      2094 AAAAGTTCCTCCCTGGAACCAATCGTTACTGCTTACATGGTTCTCTGAGAGCAATG 2153
QY      820 aHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIleProTyrA
Db      2154 TCCATTTTGACATACACAGATATTAATCTAGATTTTATGAGAGGCTGGAAACCATATG 2213
QY      840 sPheGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisT
Db      2214 ATTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATTCCTGAAATCGGAGAACATT 2273
QY      860 YrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeu
Db      2274 ATGAACGTGATCTTTTGCACTACCTTCAAGAAACCTTGGATCAGTATGCTGCTTAA 2333
QY      880 YrValIle 882
Db      2334 AAGTGATA 2341

RESULT 10
US-09-976-674-4
; Sequence 4, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2617
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-4
Alignment Scores:
Pred. No.: 0 Length: 2617
Score: 2870.00 Matches: 517
Percent Similarity: 77.50% Conservative: 134
Bec Local Similarity: 61.55% Mismatches: 187
Query Match: 61.06% Indels: 2
DB: 10 Gaps: 2
US-10-070-464-1 (1-882) x US-09-976-674-4 (1-2617)
QY      35 PheTyrValGluArgTyrSerTyrPheGlnLeuLeuValLeuLeuAlaAspThrArgIys
Db      80 TTCCAGGTGTCAGAAAGACACTGTGGAGCGGGCTCCGAGCATCATCCACGGCAGCCGCAAG 139
QY      55 TyrHisGlyTyrMetMetAlaIleValAlaProHisAspPheMetPheValIleAspArgAsnA
Db      140 TACTGGGCTCTATCTTCAACAAAGGCGCCCGACGACTTCCAGTTGTGCGAAGAGACGAGAT 199
QY      75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGlu
Db      200 GAGTCTGGGCGCCACCTCCACCGGCTTACTACTGTGGAAATCCATATATGACGCGGAGAG 259

```

OY	95	AsnThrIleuPheTYrSerGluIleProLysThrIleAsnArgAlaIleValLeuMetIleu	114
Db	260	AACTCCCTCTCTACTCTGAGATTCCAAAGAAAGTCCGAAAGAGGCTCTGCTGCTTGG	319
OY	115	SerTrpLysProIleuLeuAspLeuPheGlnAlaThrIleAspTYrGlyMetTyrSerArg	134
Db	320	TCCTGGAAGCAGATGCTGGATCATTTCCAGGCCAGCCCCACCATGGGGGTACTCTCGG	379
OY	135	GluGluGluIleuLeuArgGluArgGlySerGlyIleGlyThrValGlyIleAlaSerTyrAsp	154
Db	380	GAGGAGGAGCTGCTGAGGAGGAGGAAAGCCCTGGGGGATCTTCGGATCACTCCCAACAC	439
OY	155	TyrHisGlnGlySerGlyTYrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys	174
Db	440	TTCCACACAGAGAGTGGCTCTCTCTCTTCCAGGCCACACACAGCCTCTTCACACTGTGGC	499
OY	175	AspGlyGlyProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSer	194
Db	500	GACGGCGGCAAGAACGGCTTCATGCTGTCCTCTATGAACCGCTGAATATCAAGCCACG	559
OY	195	CysProAsnIleArgMetAspProLysLeuCysProAlaAspProAspTrpIleAlaPhe	214
Db	560	TGCTTAGGGCCCCGGATGAGCCCAAAATTCGCTTCGACAGCCCTGCTCTTCTCTTC	619
OY	215	IleHisSerAsnAspIleTyrIleSerAsnIleValIleArgGluGluArgThrLeuThr	234
Db	620	ATCATATAACAGGACCTGTGGGTGGCCAACATCGACACAGCGGAGAGACGGCGCTGACC	679
OY	235	TyrValHisAsnGluLeuAlaAsnMetGluIleAspAlaArgSerAlaGlyValAlaThr	254
Db	680	TTCTCCCAACCAAGGTTTATCCAAATTCCTGTGAATGCCCAAGTCTCGGGTGTGGCCACC	739
OY	255	PheValLeuGlnGluGluIlePheAspArgTYrSerGlyTYrTrpTrpCysProLysAlaGlu	274
Db	740	TTTCGTCATACAGAAAGATTGCAACCGCTTCACTGGGTACTGGTGTGCCCCACAGCTTCC	799
OY	275	ThrThrProSerGlyGly--LysIleLeuArgIleLeuTYrGluGluAsnAspGluSer	293
Db	800	TGGGAAGGTTCAAGAGGGCTTCMAAGCGCTGCCAATCTGTAAGAGAACTCGATAGATGCC	859
OY	294	GluValGluGluIleHisValThrSerProMetLeuGluIleThrArgArgAlaAspSerPhe	313
Db	860	GAGGTGAGGTATTCACGTCCTCTCTCTGCTGATGAAGAAAGAAACGACACTCGATT	919
OY	314	ArgTYrProLysThrGlyTYrThrIleAsnProLysValIleThrPheLysMetSerGluIleMet	333
Db	920	CGGTATCCCCAGAGACAGCGACGAAGATCCAAAGATTGGCTTGAACTGGCTGATGATCCAG	979
OY	334	IleAspAlaGluGlyArgIleIleLeuAspValIleAspLysGluLeuIleGlnProPheGlu	353
Db	980	ACTGACACACAGGGGCAAGATCGTCTTCGACCCACGAGGAAGAGAGCTGGTCAGCCCTTCACGC	1039
OY	354	IleLeuPheGluGluValGluIleTYrIleAlaArgAlaGlyTrpThrProGluGlyLysSerTYr	373
Db	1040	TCGCTGTTCCCGAAGGTGAGTGAATCCACAGGCGCCGGGTGATGCCCGGAGTGGCAAAATAC	1099
OY	374	AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValIleIleSerPro	393
Db	1100	GCTGTGGCCAAATGTTCTTGACCGCGCCACGACGATGGTCTTCAGCTGTCTCTCCCCCGG	1159
OY	394	GluLeuPheIleProValGluAspAspAlaMetGluArgGlnArgLeuIleGluIleSerVal	413
Db	1160	GCCCTGTTCATCCGACGACACAGAAATAGAGACAGCGGCTAGACCTTGCCACAGACTCTGC	1219
OY	414	ProAspSerValIleThrProLeuIleIleTYrGluGluIleThrAspIleTyrIleAsnIle	433
Db	1220	CCACAGAAATGTCACACCGATGTGTGTATACGAGGAGTGCACACAACTGTGATCAATCTT	1279
OY	434	HisAspIlePheHisValPheProGlnSerHis--GluGluGluIleGluPheIlePhe	452
Db	1280	CATGACATCTTCTATCTCTCTCCCAATCAAGAGGAGAGAGACTGTGCTTCTCTCCGC	1339

QY	453	AlaSerGluCyValysThrGlyPheAsnHisLeuTYrLysIleThrSerIleLeuYsglu	472
DB	1340	GCCATGATGATGAAAGACCGGCTTCGCCATTTGTACAAAGTACACCGCCCTTTAAATCC	139
QY	473	SerLYrLYsArgSerSerGlyYglYleuProAlaProSerAspPheLysCysProIle	492
DB	1400	CAGGGCTACGATTGGAGTGGAGCCCTTCAGCCCGGGGAAAGATTTAAATTAAGTCCCATTT	145
QY	493	LysGluGluIleAlaIleThrSerGlyGluTrpGluValLeuGlyAlaArgHisGlySerAsn	512
DB	1460	AAGGAAGATGCTGCTCTACACGCGGTAAATGGGAGTCTTGGCCAGACGACGCTTCCAAG	151
QY	513	IleGluValAspGluValAlaArgLeuValTYrPheGluGlyThrLysAspSerProLeu	532
DB	1520	ATCTGGGTCATGAGAGAAACCAAGGTGTGACTTCCAGGGCACCAAGACACGCCGCTG	157
QY	533	GluHisIleLeuTYrValValSerTYrValAsnProGlyGluValThrArgLeuThrAsp	552
DB	1580	GAGCACCACTCTACGTGTGACGCTATAGCGCGCGCGCGAGATGTCAGCTCTCACAG	163
QY	553	ArgGlyTYrSerHisAspSerCysCysIleSerGlnHisCysAspPhePheIleSerLYsTYr	572
DB	1640	CCGGGCTCTCCATAGCTGCTCCATGAGCCAGAACTTCGATGTGCTCAGCCACTAC	169
QY	573	SerAsnGlnLYsAsnProHisCysValSerLeuTYrLYsLeuSerSerProGluAspAsp	592
DB	1700	AGCAGCGTGAAGACGCGCGCTGGGTGACGTCATCAAGCTGACGCGCCCGACGACGAC	175
QY	593	ProThrCysValysThrLYsGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro	612
DB	1760	CCCCGTGCACAAAGACGCCCCGCTTCGGGCTGACAGTAAAGAGGACGACGCTGCCCGCC	181
QY	613	AspTYrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTYrGly	632
DB	1820	GATTATGTCTCTCAGAGATCTTCATTTCCACAGCGCTCGAGATGTGGCTCTACGGC	187
QY	633	MetLeuTYrLYsProHisAspLeuGlnProGlyLYsLYsTYrProThrValLeuPheIle	652
DB	1880	ATGATCTTCAAGACCCCAAGCCCTTGACGACAGGAGAAAGACACCCACCGTCTCTTTGTA	193
QY	653	TYrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLYsGlyValLYsTYrPheArg	672
DB	1940	TATGGAGGCCCCACAGTGCAGCTGGTGAATATCTCTTCAAGGATCAAGTACTTGCGG	199
QY	673	LeuAsnThrLeuAlaSerIleGlyTYrValValValIleAspAsnArgGlySerCys	692
DB	2000	CTCAACACATCGGCTCCTCGGGGTACGCGGTGTGATGACGGACGGGCTCTGCT	205
QY	693	HisArgGlyLeuLYsPheGluGlyAlaPheLYsTYrLYsMetGlyGlnIleGluIleAsp	712
DB	2060	CAGCGAGGGCTCGGTTGCAAGGGGCGCTGAAACCAATGGCGACGTGAGATTCGAG	211
QY	713	AspGlnValGluGlyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeuAspArg	732
DB	2120	GACCAAGTGGAGGGGCTGCATCTGCTGGGCCAGAAAGTATGGCTTCATGACCTGAGCCGA	217
QY	733	ValGlyIleHisGlyTYrPheSerTYrGlyGlyTYrLeuSerLeuMetAlaLeuMetGlnArg	752
DB	2180	GTTCCGATCCATGGCTGTGCTCTACGGGGGCTCTCTCTGCTCATGGGGCTAATCCACAG	223
QY	753	SerAspIlePheAspValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTYrAsp	772
DB	2240	CCCCAAGGTTCAAAGTGGCCATGCGGGGTGCCCCGGTACCGGTGGATGGCTTACGAC	229
QY	773	ThrGlyTYrThrGluArgTYrMetGlyHisAspAspGlnAsnGluGlnGlyTYrTYrLeu	792
DB	2300	ACAGGGTACACTGAGCGCTCATGACGTCCTCGAACAACACACGACGCGCTATAGAGCG	235
QY	793	GlySerValAlaMetGlnAlaGlyLYsPheProSerGluProAsnArgLeuLeuLeuLeu	812
DB	2360	GGTTCGCGGGCCCTCCACGTGGAAAGGTGCGCCATGAGCCCAACCGCTTGTTACTTC	241
QY	813	HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu	832

```
Db      2420 CACGGCTTCTGGACGAAAAACGTGCACTTTTCCACAAACACTTCCTGCTCCCAACTG 2479
Qy      833 VALARGALIGLYLysProTyrAspLeuGlnIleTyrProGlnIluArgHisSerIleArg 852
Db      2480 ATCCGAGGAGGAGAAACCTTACACGAGATCTACAGATCCCAACGAGAGACACAGTATTCCG 2539
Qy      853 ValProGlnSerGlyGlnIleTyrGlnIleHisLeuHisTyrLeuGlnIluAanLeu 872
Db      2540 TGCCCCGAGTGGGCGACGACCTATGAAATGACAGTTGCTGCACATTCTTACAGGAATACCTC 2599

RESULT 11
US-09-976-674-28
; Sequence 28, Application US/09976674
; Patent No. US2002015843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinaanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 4219
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-28

Alignment Scores:
Pred. No.: 0 Length: 4219
Score: 2870.00 Matches: 517
Percent Similarity: 77.50% Conservative: 134
Best Local Similarity: 61.55% Mismatches: 187
Query Match: 61.06% Gaps: 2
DB: 10

US-10-070-464-1 (1-882) x US-09-976-674-28 (1-4219)
Qy      35 PheTyrValIGluArgTyrSerTyrPsrSerGlnLeuLysLeuLeuLysLeuLysAspThrArgLys 54
Db      436 TTCAGAGTCAGAGACACTGCTGGACGGGCTCCGAGCATTCACACGCGACGCCCAAG 495
Qy      55 TyrHisGlyTyrMetMetAlaLysAlaProHisAspPheMetPheValIlyAspAanAsp 74
Db      496 TACTCGGGCTCATTTGTCAACAAGGCGGCCACGACTTCACAGTTTGTGCAAGAACGAGAT 555
Qy      75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyGlnAanArgLys 94
Db      556 GAGTCGGGGCCCCCACTCCACGCGCTCTACTACTCGGAAATGCCATATGGCGACGAGAG 615
Qy      95 AasnThrLeuPheTyrSerGlnIleProLysThrIleAanArgAlaIleValLeuMetLeu 114
Db      616 AACTCCCTCTCTACTCTAGATTCCCAAGAAAGCTCCGAAAGAGAGCTGCTGCTCCG 675
Qy      115 SerTyrLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134
Db      676 TCTCGAAGACAGATGCTGATATCTTCCAGGCCACCCACACATGGGGCTTACTCTCG 735
Qy      135 GlnGlnGlnLeuLeuArgGlnArgLysArgIleGlyThrValGlyIleAlaSerTyrAsp 154
Db      736 GAGGAGGAGACTGCTGAGGAGCGGAAACCCCTGGGGGCTTCGCGCATTCCTTACGAC 795
Qy      155 TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys 174
Db      796 TTCACAGGAGAGAGGAGCTCTTCTCTTCCAGGCCACGACAGCATCTTCCACTGCCGC 855
Qy      175 AspGlyGlyProGlnIlePheThrGlnGlnProLeuArgProAanLeuValGluThrSer 194
```

```
Db      856 GACGGGGGCAAGAACGGCTTCACTGCTGCCCTATGAAACCGCTGGAATATCAAGACCAG 915
Qy      195 CysProAanIleArgMetAspProLysLeuCysProAlaAspProAspTrpIleAlaPhe 214
Db      916 TGCTCAGGGGCCCGGATGAGACCCCAAAATCTGCCCTGCGACACCTGCTCTTCTCCTTC 975
Qy      215 IleHisSerAspAspIleTrpIleSerAanIleValThrArgGlnGlnArgGlyLeuThr 234
Db      976 ATCAATTAACAGCACTGTGGGTGGCCAAATCATGACACAGGACGAGGCGGGCGGTGAC 1035
Qy      235 TyrValHisAanGlnLeuAlaAanMetGlnGlnAspAlaArgSerAlaGlyValAlaThr 254
Db      1036 TCTCGCCACCAAGCTTATCCATATGCTTCGAGATGACCCCAAGCTCGGGGTGGCCACC 1095
Qy      255 PheValLeuGlnGlnIlePheAspArgTyrSerGlyTyrTrpCysProLysAlaGln 274
Db      1096 TTGCTATACAGGAAGATTCGACCCGCTTACACTGGGTACTGCTGCCACAGGCTCC 1155
Qy      275 ThrThrProSerGlyLys---LysIleLeuArgIleLeuTyrGlnGlnAanAspGlnSer 293
Db      1156 TGGAAAGTTTCAGAGGCTCCAGACGCTGCGAATCTGTATAGGAAGTCGATGAGTCC 1215
Qy      296 GlnValGlnIleIleHisValThrSerProMetLeuGlnThrArgArgAlaAspSerPhe 313
Db      1216 GAGGTGAGGTATTCATCAGTCCCTCTCTGCGCTGAAGAAAGAAAGACGACTCTGAT 1275
Qy      314 ArgTyrProLysThrGlyThrAlaAanProLysValThrPheLysMetSerGlnIleMet 333
Db      1276 CGGTACCCGACGACGACGACGAAGATCCCAAGATTGCTTGAATCGGCTGAGTTCAC 1335
Qy      334 IleAspAlaGlnIleArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGln 353
Db      1336 ACTGACAGCGAGGCAAGATGCTGACCCAGCAAGAAAGAGCTGTGCGAGCCCTTACGC 1395
Qy      354 IleLeuPheGlnGlnIleValGlnTyrIleAlaArgAlaGlyTrpThrProGlnGlyLysTyr 373
Db      1396 TCGCTTCTCCGAAGTGAAGTATGCTGCGAGGCGGGGTGACCCGGATGGCAATAC 1455
Qy      374 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393
Db      1456 GCGTGGCCATGTTCTGTGACCGGCCACGAGTGGCTCAGCTGCTCTCCGCCCGC 1515
Qy      394 GlnLeuPheIleProValGlnAspAspValMetGlnArgGlnArgLeuIleGlnSerVal 413
Db      1516 GCCCTGTATCCCGACACAGAAATGAGAGACGCGCTACCTCTGCCAGACGCTC 1575
Qy      414 ProAspSerValThrProLeuIleIleTyrGlnGlnThrThrAspIleTrpIleAanIle 433
Db      1576 CCAGAGAAATGTCAGCCGATATGCTGATGAGAGAGTCAACCACTTGATCATGATGTT 1635
Qy      434 HisAspIlePheHisValPheProGlnSerHis--GlnGlnGlnIleGlnPheIlePhe 452
Db      1636 CATGACATCTCTATCTCTTCCCAATCAGAGGAGAGAGACAGACGCTGCTTCCGCG 1695
Qy      453 AlaSerGlnCysLysThrGlnIlePheArgHisIleLeuTyrLysIleThrSerIleLeuLysGln 472
Db      1696 GCCAATGAATGCAAGACGCGCTTTCGCATTGTATGAAGATCCGCCCGTTTAAATCC 1755
Qy      473 SerLysTyrLysArgSerSerGlyLysProAlaProSerAspPheLysCysProIle 492
Db      1756 CAGGGCTTACATGAGTACGCTTCAAGCCCTTCAAGCCCGGGAAGATTAATTAAGTCCCAAT 1815
Qy      493 LysGlnGlnIleAlaIleThrSerGlnGlnTyrGlnValIleGlnIlyArgHisGlySerAan 512
Db      1816 AAGGAAGAGATGCTCTGACCAAGCGGTGAATGGAGATTGGCCAGGACCGCTCCAG 1875
Qy      513 IleGlnValAspGlnValArgArgLeuValTyrPheGlnGlnIlyThrLysAspSerProLeu 532
Db      1876 ATCTGGGTCAATGAAGAGACCAAGCTGTGATCTTCCAGGGACCAAGAGACACGCGCTG 1935
Qy      533 GlnHisIleLeuTyrValIleSerTyrValAanProGlnGlnValThrArgLeuThrAsp 552
```

Db 1936 GAGCACCACCTTACGTGTCAGCTATAGAGCGCGCCGACATCGATCCCTCACCACG 1995  
Qy 553 ArgGlyTyrSerHisSerCysGlyLeuSerGlnHisCysAspPhePheLeuSerLysTyr 572  
Db 1996 CCCGGCTTCTCCATAGCTCTCTCCATAGCCAGAACTTCACATGTTCTGTCAGCACCTAC 2055  
Qy 573 SerAsnGlnLysAspProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 592  
Db 2056 AGCAGCGTGAACACCGCGCTGCTGTCACAGCTCACAAGCTGAGCGCGCCCGACAGACG 2115  
Qy 593 ProThrCysLysThrLysGluPheTyrAlaThrLleLeuAspSerAlaGlyProLeuPro 612  
Db 2116 CCCCTGCACAAAGACGCCCGCTTCTGCGCTAGATGATGAGGACGACCTGCCCCCG 2175  
Qy 613 AspTyrThrProProGluLlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 632  
Db 2176 GATTATGTTCTCTCAGAGATCTTCATTCCACACGCGCTCGAGTGTGCGGCTCTACGCG 2235  
Qy 633 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLleuPhe1le 652  
Db 2236 ATGATCTACAAAGCCCAACGCGCTTGACCCAGGAGAAAGACCCCAACCGTCTCTTTGTA 2295  
Qy 653 TyrGlyGlyProGluValGlnLeuValAspAsnAspPheLysGlyValLysTyrPheAsp 672  
Db 2296 TATGAGGCGCCCAAGGTGACGTGTGTAATACCTCTTCAAAAGCATCAAGTACTTGCGG 2355  
Qy 673 LeuAsnThrLeuLysSerLeuGlyTyrValValValLleAspAsnAspGlySerCys 692  
Db 2356 CTCACACACTGCGCTTCTCTGAGGCTACCGCGGTGTGTGATGACGCGACAGGCGCTCTGT 2415  
Qy 693 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnLleGluLleAsp 712  
Db 2416 CACGAGGCGCTTCTGTAAGGCGCCCTGAAAAAACAAATGGGCCAGGTGAGATCGAG 2475  
Qy 713 AsnGlnValGlnGlyLeuGlnTyrLleuLysSerArgTyrAspPheLleAspLeuAspArg 732  
Db 2476 GACCAAGTGAAGGAGCGCTGCGAGTTCGTGCGAGAAAGATGCTTCATCGACCTGAGCGGA 2535  
Qy 733 ValGlyLleHisGlyTyrPheSerTyrGlyGlyTyrLleuSerLeuMetAlaLeuMetGlnArg 752  
Db 2536 GTTGCCATCGATGGCTGTGCTCTACGCGGCGCTTCTCTCGCTCATGGGGCTTATTCACAG 2595  
Qy 753 SerAspLlePheAspValAlaLleAlaGlyAlaProValThrLeuTyrLlePheTyrAsp 772  
Db 2596 CCCCAAGTGTTCAAAGGTGCGCATCGCGGTCGCCCGGTCAACGCTGAGTGGCTACGAC 2655  
Qy 773 ThrGlyTyrThrLysTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeu 792  
Db 2656 ACAGGCTACCTAGCGGCTACATGAGACGCTCCGAGAAACACGACGCGCTATGAGGCG 2715  
Qy 793 GlySerValAlaMetGlnAlaGlnLysPheProSerGluProAsnArgLeuLeuLeu 812  
Db 2716 GGTTCCTGCGCTGCGACGTCGAGAAAGCTGCCCAATAGCCCAACCGCTTATTCCTC 2775  
Qy 813 HisGlyPheLeuAspGlnValHisPheAlaHisThrSerLleLeuLeuSerPheLeu 832  
Db 2776 CACGGCTTCTGACCAAAACGTGCACTTTTCCACAAACTTCCGCTCCCAACTG 2835  
Qy 833 ValArgAlaGlyLysProTyrAspLeuGlnLleTyrProGlnLysArgHisSerLleArg 852  
Db 2836 ATCCGAGCAGGGAACCTTACAGCTTCAGATCTACCCCAACGAGAGACACATATTTCG 2895  
Qy 853 ValProGluSerGlyGlnHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnValAsn 872  
Db 2896 TGGCCGAGTCCGGCGGACGCTATGAAAGTCAACGTTGTGCACTTCTACAGGAATACCTC 2955

RESULT 12  
US-09-976-674-24

; Sequence 24, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akinsanya, Karen

; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976, 674  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24  
; LENGTH: 4302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-976-674-24  
Alignment Scores:  
Pred. No.: 0 Length: 4302  
Score: 2870.00 Matches: 517  
Percent Similarity: 77.50% Conservative: 134  
Best Local Similarity: 61.55% Mismatches: 187  
Query Match: 61.06% Indels: 2  
DB: 10 Gaps: 2  
US-10-070-464-1 (1-882) x US-09-976-674-24 (1-4302)  
Qy 35 PheTyrValGlnArgTyrSerTyrPheSerGlnLeuLysLeuLleAspThrArgLys 54  
Db 436 TTCACAGTGAAGACCTGTGTGAGCGGCTCCGAGCATCTCCACGCGACCGCCAG 495  
Qy 55 TyrHisGlyTyrMetMetAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74  
Db 496 TACTCGGCTCATTTGTCAAAAGCGCCCGACGACTTCAGTGTGTGCGAAGACGAT 555  
Qy 75 ProAspGlyProHisSerAspArgLleTyrTyrLeuAlaMetSerGlyGluAsnArgGlu 94  
Db 556 GAGTCTGGGCGCCACCTCCACCGCTCTACTACCTGGAAATGCATATGGCAGCCGAG 615  
Qy 95 AsnThrLeuPheTyrSerGlnLleProLysThrLleAsnArgAlaAlaValLleuMetLeu 114  
Db 616 AACTCCCTCTCTACTGTGATTTCCAAAGAGTCCGGAAGAGCTGCTGCTCTG 675  
Qy 115 SerTyrLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134  
Db 676 TCTTGAAACAGATGCTGATGATCTTCCAGGCCACCGCCACATGGGCTTACTCGG 735  
Qy 135 GlnGlnGlnLeuLeuArgLysArgLleGlyThrValGlyLleAlaSerTyrAsp 154  
Db 736 GAGGAGAGCTGCTGAGGAGCGGAAACCGCTGGGGGTCTTCGGCATCACCTCTAGAC 795  
Qy 155 TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyLleTyrHisValLys 174  
Db 796 TTCACAGCGAAGAGTGGCTTCTCTTCCAGGCGCAGAACGACTTCCACTCGC 855  
Qy 175 AspGlyGlyProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGlnThrSer 194  
Db 856 GACGGGCGAAGAACGCTTCAATGCTGCCCTATGAACCGCTGGAATATCAAGACCG 915  
Qy 195 CysProAsnLleArgMetAspProLysLeuCysProAlaAspProAspTyrLleAlaPhe 214  
Db 916 TGTCTAGGCGCCCGGATGAGACCCCAAAATCTGCGCGACCTGCTCTCTCTC 975  
Qy 215 LleHisSerAsnAspLleTyrLleSerAsnLleValThrArgGlnGlnArgGlnLeuThr 234  
Db 976 ATCAATTAACAGGACCTGTGTGGGCAATTCGACAGCGCGAGGAGCGCGCTGACC 1035  
Qy 235 TyrValHisAsnGlnLeuLleAsnMetGlnGluAspAlaArgSerAlaGlyValAlaThr 254  
Db 1036 TTCTGCCACCAAGGTTATTCATATGTCCTGATGACCCCAAGTCTCGGGTGTGGCCACC 1095  
Qy 255 PheValLleGlnGlnGlnLysPheAspArgTyrSerGlyTyrTyrTyrCysProLysAlaGlu 274  
Db 1096 TTCGTATACAGAAAGATTGACCGCTTCACTGGGTACTGTGTGTCGCCACAGCTCC 1155

```

Oy 275 ThrThrProSerGlyIle---LysIleLeuArgIleLeuTyrGluGluLeuAspGluSer 293
Db 1156 TGGGAAGTTCAGAGAGCCCTCAAGACGCTCCGAATCTGTAATAGAAAGCATGTGATGCC 1215
Oy 294 GluValGluIleIleHisValThrSerProMetLeuGluThrArgAlaAspSerPhe 313
Db 1216 GAGGTGAAGGTCTATCAGCTCCCTCTCTGCGCTAGAAAGAAAGAAAGACGATCTGTAT 1275
Oy 314 ArgTyrProLeuSerThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMet 333
Db 1276 CGGTACCCAGGACAGACAGCAAGAAATCCCAAGATTCCTTGAACATGGCTGTGATCCAG 1335
Oy 334 IleAspAlaGluGlyValArgIleIleAspValIleAspLeuGluLeuIleGlnProPheGlu 353
Db 1336 ACTGACAGCGAGGCAAGATCTCTGACCCAGAGAAAGAGCTGTGTGACGCCCTTCAGC 1395
Oy 354 IleLeuPheGluGlyValGluTyrIleAlaArgAlaGlyTyrPheProGluGlyLysTyr 373
Db 1396 TCGCTGTCCCGAAGGTGAGTACATGCGCCAGGGCGGGTGGACCGGGATGGCAATAC 1455
Oy 374 AlaTyrSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValIleuIleSerPro 393
Db 1456 GCTTGAGGCAATGCTCTGACCGGCGCCAGACAGTGGCTCAGCTCTCTCCCTCCCGCG 1515
Oy 394 GluLeuPheIleProValGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413
Db 1516 GCCCTGTATCCCGAGCAACAGATGAGAGACGGGTAGCCCTCTCCAGAGCTGTTC 1575
Oy 414 ProAspSerValThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIle 433
Db 1576 CCAGAGAAATGTCAGCGCTATGTGTGTCAGAGAGTCAACCACTGATCATATGTT 1635
Oy 434 HisAspIlePheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePhe 452
Db 1636 CATTGACATTTCTATCTCTCCCTCCCAATCAGAGAGAGACGACGATCTGCTTCTCCCGC 1695
Oy 453 AlaSerGluCysLeuThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGlu 472
Db 1696 GCCAATGATGCAAGACGGGCTTCTGCTCCATTGTACAAAGTCAACGCCCTTTAAATCC 1755
Oy 473 SerLysTyrLysArgSerSerGlyLysProAlaProSerAspPheLysCysProIle 492
Db 1756 CAGGCTCAGATTGGGTGAGGCTTCAACCGCGGGGAGATGAATATTAAATGAGCCCAT 1815
Oy 493 LysGluGluIleAlaIleThrSerGlyGluTyrGluValLeuGluValArgHisGlySerAsn 512
Db 1816 AAGGAAGATTTGCTCTGACGAGCGGTGATGGAAGTTTGTGCGGACGACGCTTCCAG 1875
Oy 513 IleGlnValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 532
Db 1876 ATCTGGGTCAATGAGAGAGACCAAGCGTGTACTTCCACAGGACCAAGACACGCCGCTG 1935
Oy 533 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 552
Db 1936 GAGCAACCACTTACGTGTGATGAGCGGCGCGGAGATGTGATGAGCTTCAACACAG 1995
Oy 553 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyr 572
Db 1996 CCCGCTTCTCCATAGCTGCTGCTCCATGACCCCAAACTTGACATGTCTCCACCATAC 2055
Oy 573 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 592
Db 2056 AGCAGCGTAGAGACCGCGCGCTGCGACGCTCAAGCTAAGCTAGAGGGCGCCGACAGCAG 2115
Oy 593 ProThrCysLysThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuPro 612
Db 2116 CCCCTGCAACAGACCGCGCTTCTGGGCTAGACATGATGAGAGCAGCAGCTGCCCGCG 2175
Oy 613 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrThrLeuTyrGly 632
Db 2176 GATTATGTTCTCTCCAGATCTTCCATTTCACACGCGCTGGATGTGCGGCTTCAACGCG 2235

```

```

Oy 633 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 652
Db 2236 ATGATCTTCAAGGCCCAACGCCCTTGGACGAGGAGAAAGACACCCACGCTCTTGTGA 2295
Oy 653 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 672
Db 2296 TATGAGAGCCCAAGAGTGGAGCTGTGAAATTAATCTTCAAGAGCATCAAGTACTTGGCG 2355
Oy 673 LeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCys 692
Db 2356 CTCACACACTGGCTCCCTCGGCTTACGCGGTGTGTGATGACCGCAGGGGCTCTCGT 2415
Oy 693 HisArgGlyLeuLeuPheGluGluValAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 712
Db 2416 CAGCGAGGCTTCCGTTCCAGAGGCTCTGAAAGAACCAAAATGGAGGAGGATGACAG 2475
Oy 713 AspGlnValGluGlyLeuGlnIleValAlaSerArgTyrAspPheIleAspLeuAspArg 732
Db 2476 GACCAAGTGAAGGCTTGGACGATTCGTGGCCGAGAAATATGGCTTATCCAGCTTGAGCCGA 2535
Oy 733 ValGlyIleHisGlyTyrPheTyrGlyTyrLysSerLeuMetAlaLeuMetGlnArg 752
Db 2536 GTTGCATTCATGCGCTGCTTCAAGGGGCTTCTCTCCCTCATGGGGCTAATCCACAG 2595
Oy 753 SerAspIlePheArgValAlaIleAlaGlyValAlaProValThrLeuTrpIlePheTyrAsp 772
Db 2596 CCCAGGTTCATAGGTGACATCCGCGGTGCGCGGTGACCGCTCGATGAGCTTACGAC 2655
Oy 773 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeu 792
Db 2656 ACAGGGTACACTGAGAGGCTTACATGACGCTCCGAGAACACACAGACCGCTATGAGCGC 2715
Oy 793 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 812
Db 2716 GGTTCGTCGCTCGCTGACCTGAGAGAAAGCTGACCAATGAGCCCAACCGCTTGTATCTTC 2775
Oy 813 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
Db 2776 CACGGCTTCTGAGAGAAACGTCATTTTCCACACAACTTCTCGTCTTCCAACTG 2835
Oy 833 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 852
Db 2836 ATCCGAGCAGGAGAACTTACCAAGCTCCAGATCTACCCCAAGAGAGACACAGTATTCGC 2895
Oy 853 ValProGluSerGlyGluHisTyrGluLeuHisLysLeuHisTyrLeuGlnGluAsnLeu 872
Db 2896 TCCCGCGAGTCCGGCGAGCAGCATATGAAATGACGTTGCTCAGTCTTCTACAGGAATACCTC 2955

```

## RESULT 13

```

US-09-976-674-36
/ Sequence 36, Application US/09976674
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Qi, Steve
/ APPLICANT: Akinsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Junien, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
/ FILE REFERENCE: 70669
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 36
/ LENGTH: 4180
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-976-674-36

```

Alignment Scores: 0 Length: 4180

Pred. No.: 0

Score: 2820.50 Matches: 510  
 Percent Similarity: 76.43% Conservative: 132  
 Best Local Similarity: 60.71% Mismatches: 183  
 Query Match: 60.01% Indels: 15  
 DB: 10 Gaps: 3

US-10-070-464-1 (1-882) x US-09-976-674-36 (1-4180)

Qy PheTyValGluArgTyrSerGlnLeuLysLeuAlaAspThrArgLys 54  
 Db TTTCCAGTGCAGAGCACTCGTGGACGGGCTCCGAGCATCATCCAGCGACGGCAAG 495  
 Qy TyrHisGlyTyrMetAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74  
 Db TACTCGGGCCCTCATTTGTCACAGCGCGCCACGCTTCCAGTTTGGCACAAGACGGAT 555  
 Qy ProAspGlyProHisAspArgLysLeuTyrTyrLeuAlaMetSerGlyLysAsnArgLys 94  
 Db GAGTCTGGGCCCCCTCCACCGCTCTACTACTGGAATGCCATATGGCAGCGCGAG 615  
 Qy AsnThrLeuPheTyrSerGlnLeuProLysThrIleAsnArgAlaAlaValIleuMetLeu 114  
 Db AACCTCCCTCTCTACTCTGAGATTCACAGAAAGTCCGAAAGAGCTCTGCTCTCTG 675  
 Qy SerTyrLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134  
 Db TCTCGAAGACGATGCTGATCATTTCCAGGCCACGCCCATGGGGTCTACTCTCGG 735  
 Qy GlnGlnGlnLeuLeuArgGlnLysArgLysLeuGlyThrValGlyIleAlaSerTyrAsp 154  
 Db GAGAGAGAGCTGCTGAGGAGCGGAAACGCTGGGGCTTTCGGCATCACCTCTACGAC 795  
 Qy TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys 174  
 Db TTCCACAGCCAGATGGCTCTTCTCTCCAGGCCACAGAACACCTCTTCCATGCGCG 855  
 Qy AspGlyGlyProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194  
 Db GACGCGCGCAAGACGGCTTCATGCTGCTGTAAGAACCGCTGAAATCAAGACCCAG 915  
 Qy CysProAsnIleArgMetAspProLysLeuCysProAlaAspProAspTyrIleAlaPhe 214  
 Db TGCTCAGGGCCCCGGAATGAGCCCAAAATCTGCCCTGCGACCTTCTTCTCTCTC 975  
 Qy IleHisSerAsnAspIleTyrPheSerAsnIleValThrArgGlnGluArgLysLeuThr 234  
 Db ATCAATAACAGCGACCTGTGGGTGCCAATCCAGACAGCGCGAGCGCGCTACCC 1035  
 Qy TyrValHisAsnGlnLeuAlaAsnMetGlnGluAspAlaArgSerAlaGlyValAlaThr 254  
 Db TTCTGCCACCAAGGATTATCCAAATGCTCTGGATGACCCCAAGTCTGGGGGTGTGGCACC 1095  
 Qy PheValIleuGlnGlnGlnIlePheAspArgTyrSerGlyTyrTyrProLysValGlu 274  
 Db TTCTCTATACAGAAAGTTCAGCCGCTTCACTGGGTACGTGGGTGCTCCACAGCTCC 1155  
 Qy ThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlnGluAsnAspGluSer 293  
 Db TGGGAAGCTTCAGAGGGCTCAAGACCGTGGGAATCTGTATAGGAAGTGAATGCC 1215  
 Qy GluValGlnIleIleHisValThrSerProMetLeuGluThrArgAlaAspSerPhe 313  
 Db GAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAAGAAAGAAAGAGCGAGCTCTGAT 1275  
 Qy ArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMet 333  
 Db CGGTACCCCAAGACAGCGACAGAAATCCCAAGATTGCTTGAACCTGGCTGACTTCCAG 1335  
 Qy IleAspAlaGlnGlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGlu 353  
 Db ACTGACAGCGACGAGCATGCTCCAGCCAGAGAAAGAGCTGTGGCAGCCCTTCCAG 1395  
 Qy IleLeuPheGlnGluValGluTyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyr 373

Db TCGCTGTTCGGAAGTGAAGTACATGCCAGCGCCGGGTGAGACCGGGAATGGCAATAC 1455  
 Qy AlaTyrSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValIleuIleSerPro 393  
 Db GCGTGGGCGCATGTTCTCTGAGCCGCGCCACAGTGTCTCCAGCTCTCTCTCCCGCG 1515  
 Qy GluLeuPheIleProValGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413  
 Db GCCCTGTTCATCCGACACAGAAATGAGAGAGCGGCTAGGCTCTGCCAGAGCTGTTC 1575  
 Qy ProAspSerValThrProLeuIleIleTyrGlnGluThrThrAspIleTyrPheAsnIle 433  
 Db CCGAGAAATGCCAGCGGATGTGTGTACAGAGAGGTCAACACGCTGGATCAATGTT 1635  
 Qy HisAspIlePheHisValPheProGlnSerHisGlnGlnGlnIleGluPheIlePhe 452  
 Db CATACATCTTCTATCTCTTCCCAATCAGAGAGAGACGAGCTCTGCTTCTCTCGC 1695  
 Qy AlaSerGlnCysLysThrGlyPheArgHisLeuTyrIleValIleThrSerIleLeuLysGlu 472  
 Db GCCAATATGCAAGACCGGCTCTGCTGATTTGTACAAAGTACACCGCTTTTAAATCC 1755  
 Qy SerLysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIle 492  
 Db CAGGCTACGATTTGAGATGAGCCCTTCCAGCCCGGGAAGATTAATTAATGCTCCATT 1815  
 Qy LysGlnGlnIleAlaIleThrSerGlyGlnTyrGlnValLeuGlyArgHisGlySerAsn 512  
 Db AAGGAAAGATGCTGTGACACCGGTGAATGGAGGTTTGGGAGGACCGGCTCC--- 1872  
 Qy IleGlnValAspGluValArgArgLeuValTyrPheGlnTyrLysAspSerProLeu 532  
 Db -----AAGGACACCAAGACACCCCTG 1896  
 Qy GluHisIleLeuTyrValIleSerTyrValAsnProGlyGluValThrArgLeuThrAsp 552  
 Db GAGCACCACTTACGTCGTGACGTAAGAGCGCGCGGAGATCGAACGCTCACCAGC 1956  
 Qy ArgGlyTyrSerHisSerCysIleSerGlnHisCysAspPhePheIleSerLysTyr 572  
 Db CCGGCTTCTCCATAGCTCTCATAGGCAAACTTCACATGTTCGTGACGACTAC 2016  
 Qy SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 592  
 Db AGCAGCTGAGCAGCGCCCTGCTGACGCTCAAGCTTAGCGCCGCCAGAGACAC 2076  
 Qy ProThrCysLysThrLysGlnPheTyrAlaThrIleLeuAspSerAlaGlyProLeuPro 612  
 Db CCCCTGCAAGAGCGCCCTTCTGGGTAGCATGATGAGAGCAGCCAGCTGCCCG 2136  
 Qy AspTyrThrProProGlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 632  
 Db GATTATGTTCTCCAGAGATCTTCATTTCCACACCGGCTCGAGTGTGCGGCTCTACGGC 2196  
 Qy MetLeuTyrIleProHisAspLeuGlnProGlyLysLysTyrProThrValIleuPheIle 652  
 Db ATGATCTACAAAGCCCAACGCTTGCAGCCAGGAAAGAACACCCACCTCTCTTGTGA 2256  
 Qy TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 672  
 Db TATGAGAGGCCCAAGGTGACGTGTGAATTAATCTCTTCAAGAGCATCAAGTACTGGCG 2316  
 Qy LeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgLysSerCys 692  
 Db CTCAACACACTGGCTCTCTGAGGCTACGCCGTGTGTGTTTACGCGACAGGGCTCTCGT 2376  
 Qy HisArgGlyLeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGlnIleAsp 712  
 Db CAGCGAGGGCTGTGTTTCAAGAGGCGCTGAAACCAATATGGCCAGGTGAGATCGAG 2436  
 Qy AspGlnValGlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 732

[illegible]

```

RESULT 14
; Sequence 34, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 4263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-34

Alignment Scores:
Pred. No.: 0
Score: 2820.50
Percent Similarity: 76.43%
Best Local Similarity: 60.71%
Query Match: 60.01%
DB: 10
Gaps: 3
US-10-070-464-1 (1-882) x US-09-976-674-34 (1-4263)

Qy 35 PheYrYalGluAgtYrSerTIpSerGlnLeulYslySleuLeuAlaApThrArGlyS 54
Db 436 TTCAGGATGCAAGACACTCGTGGACGGCTCCGCGACATTCACCGCACCGCCAG 495
Qy 55 TyrHisGlyTrwCmetAlaYsAlaProHisApPheMetPheValYsAsGanaAsp 74
Db 496 TACCGGGCCCTCATTCATCAACAAGGCCCCACGACTTCAGATTGTGCGAAGACCGAT 555
Qy 75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyIuAsnArgIu 94

```

[illegible]







QY 175 AspGlyValProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194  
 DB 856 GACGGGGGCAAGAACGGCTTCATGGTGTCCCTATGAAACCGCGAATATCAAGACCCAG 915  
 QY 195 CysProAsn1LeuArgMetAspProGlyLeuGlyProAlaAspProAspTrp1LeuAlaPhe 214  
 DB 916 TGTCTACGGGGCCCGGATGAGACCCCAAAATCTGCTCCGACCCCTGCTCTCTCTCC 975  
 QY 215 IleHisSerAsnAsp1LeuTrp1LeuSerAsn1LeuValThrArgGluGluArgArgLeuThr 234  
 DB 976 ATCAATACAGCAACCTGTGGTGGCCCAATCGACAGCAGGCGAGAGCGGCGGTAC 1035  
 QY 235 TyrValHisAsnGluLeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThr 254  
 DB 1036 TTCTGCACCAAGATTATCAATGTCTCGATGACCCCAAGTCTGGGGGTGGCCAC 1095  
 QY 255 PheValLeuGlnGluGluPheAspArgTyrSerGlyTyrTrpTrpCysProGlyAlaGlu 274  
 DB 1096 TTGCTCATACAGAGAGTTCGACCGCTTCACTGGGTACTGGTGGTCCCCACAGCTCC 1155  
 QY 275 ThrThrProSerGlyGly--LysIleLeuArgIleLeuTyrGluGluAsnAspGluSer 293  
 DB 1156 TGGGAAGGTTCAAGGGCTTCAGAGCCCTCGAATCTGTATGAGAGATCGATGAGTCC 1215  
 QY 294 GluValGluLeuIleHisValThrSerProMetLeuGluThrArgArgAlaAspSerPhe 313  
 DB 1216 GAGGTGAGGTCAATTACGTCCTCCCTCTCTGCGCTAGAAAGAAAGAACGAGCTGTAT 1275  
 QY 314 ArgTyrProGlyThrGlyThrAlaAsnProGlyValThrPheGlyMetSerGlu1LeuMet 333  
 DB 1276 CGGTACCCCGAGCAGGACGACCAAGAAATCCCAAGATTCCTTGAAGCGCTGAGTTCAG 1335  
 QY 334 IleAspAlaGluGluArgIleIleAspValIleAspGlyLeuLeuIleGlnProPheGlu 353  
 DB 1336 ACTGACAGCCAGGCGCAAGATCTCTCGACCCAGAGAGAGAGCTGTGGCAGCCCTTAC 1395  
 QY 354 IleLeuPheGluGluValGluTyrIleAlaArgAlaGlyTrpThrProGluGlyLeuTyr 373  
 DB 1396 TCGCTGTCCCGAAGGCGAGTACATGCGCAGGGCGCGGTGAGACCCGCGATGCAATAC 1455  
 QY 374 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393  
 DB 1456 GCGTGGCCCATGTTCTCGACCGGCGCCACAGCTGCTCAGCTGCTCCCTCCCGCC 1515  
 QY 394 GluLeuPheIleProValGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413  
 DB 1516 GCCCTGTTCATCCCGAGCAGACAGAAATGAGAGACAGCGGCTACCTTCCCAAGCTGC 1575  
 QY 414 ProAspSerValThrProLeuIleIleTyrGluGluThrThrAspIleTrpIleAsnIle 433  
 DB 1576 CCCAGAAATGTCCAGCCGTATGTGTACGAGGAGGTACCAACGTCGTGATCAATGTT 1635  
 QY 434 HisAsp1LeuPheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePhe 452  
 DB 1636 CATGACATCTTCATCTTCCTTCCCAATCGAGGAGAGAGAGAGCTGCTCTTCCCGC 1695  
 QY 453 AlaSerGluCysAlaThrGlyPheArgHisLeuTyrIleThrSerIleLeuLeuValGlu 472  
 DB 1696 GCCAATGAATGCAAGACCGGCTTCTGCAATTTGTAACAAATCACCGCGTTTAAATCC 1755  
 QY 473 SerLysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIle 492  
 DB 1756 CAGGGCTACGATGAGTGAAGCCCTTCAGCCCGGGGAAAGATGAATTTAAGTCCCAT 1815  
 QY 493 LysGluGluIleLeuIleThrSerGlyGluTrpGluValLeuGluLysArgHisGlySerAsn 512  
 DB 1816 AAGGAAGAGTTGCTCTGACAGCGGGAATGGAGGTTTGGCGAGGCGACGGCTCCAG 1875  
 QY 513 IleGlnValAspGluValAlaArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 532  
 DB 1876 ATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCAGCAAGAGCAGCGCGT 1935  
 QY 533 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 552

DB 1936 GACACACACCTTACAGTGGTCAAGCTATGAGCGCGGAGATTCGTAACGCTCACCAG 1995  
 QY 553 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPheIleSerLysTyr 572  
 DB 1996 CCGGCTTCTCCATAGCTGTCTCATTAGAGCAAACTTCACATGTTCGTGACGACTAC 2055  
 QY 573 SerAsnGlnLysAsnProHisCysValSerLeuTyrIleLeuSerSerProGluAspAsp 592  
 DB 2056 AGCAGGCTAGACACCGCCCTCGTGCACGTCTTACAGCTGAGCGGCCCGCAGACGAC 2115  
 QY 593 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 612  
 DB 2116 CCCCTGACAAAGACGCCCGCTTCTGGGCTAGCATATGAGGAGCAGCAGCTGCCCCCG 2175  
 QY 613 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 632  
 DB 2176 GATTATGTTCTCCAGAGATCTTCATTTCCACAGCGGCTCGGATGTGCGGCTCTACGC 2235  
 QY 633 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 652  
 DB 2236 ATGATCTCAAGCCCTCCAGCCTTGCAGCCAGGAGAGAGACCCACCGTCTCTTTGTA 2295  
 QY 653 TyrGlyValProGlnValGluLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 672  
 DB 2296 TATGAGAGGCCCGCAGGTGACGTGTGAATATCTCTTCAAGAGCATCAAGTACTTGGCG 2355  
 QY 673 LeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCys 692  
 DB 2356 CTCACACACTGGCCCTCCGCGGCTACCGCTGTGTGTATTAACGCGCAGGGGCTCTGT 2415  
 QY 693 HisArgGlyLeuLysPheGluGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 712  
 DB 2416 CAGCGAGGCTTGGTTCGAGGGGCGCTGAAACCAATATGGCCAGCGTGGAGATCGAG 2475  
 QY 713 AspGlnValGluGluLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 732  
 DB 2476 GACCAAGTGAAGGCGCTCGCATTCGTGGCCGAGAGATAGGCTTACCTGACCTGAGCGCA 2535  
 QY 733 ValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752  
 DB 2536 GTTGCCATCATAGCTGTGTCTTACGAGGGGCTTCTCTCGTCACTGAGGGCAATCCACAG 2595  
 QY 753 SerAsp1LeuPheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAs 772  
 DB 2596 CCCAGCTGTTCAAGG----- 2611  
 QY 772 pThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLe 792  
 DB 2611 ----- 2611  
 QY 792 uGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 812  
 DB 2612 -----CCCAACCGCTTGCCTTACCT 2631  
 QY 812 uHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLe 832  
 DB 2632 CACGGCTTCTCGAGAAACGTGCACCTTTTCCACACAAATCTCTGCTCCCAACT 2691  
 QY 832 uValAlaArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 852  
 DB 2692 GATCCGAGCAGGGAACCTTACACAGCTCCAGATCTACCCCAACGAGACACAGATTTGC 2751  
 QY 852 gValProGluSerGlyGluHisTyrGluLeuHisIleLeuLeuHisTyrLeuGlnGluAsnLe 872  
 DB 2752 CTGCCCGAGTGGGGGACACTATGAATCAAGTTCGTCACCTTCTACAGAAATACCT 2811  
 QY 872 u 872  
 DB 2812 C 2812

Job time : 792.423 secs

0

0

0

0

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:24:12 ; Search time 5299 Seconds  
(without alignments)  
4045.398 Million cell updates/sec

Title: US-10-070-464-1  
Perfect score: 4700  
Sequence: 1 MAAMETQQLGVEIFETADC.....HLHYLGNTGSRALAKVI 882

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10070464/runat\_15102003\_113553\_24829/app\_query.fasta\_1.2652  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human0.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pio -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10070464 @CGN 1.1 8056 @runat\_15102003\_113553\_24829 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database :  
EST :  
1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: gb\_estc1:\*  
10: gb\_estc2:\*  
11: gb\_hic:\*  
12: gb\_estc3:\*  
13: gb\_estc4:\*  
14: gb\_estc5:\*  
15: em\_estba:\*  
16: em\_estba:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_hum:\*  
20: em\_gss\_hum:\*  
21: em\_gss\_hum:\*  
22: em\_gss\_hum:\*  
23: em\_gss\_hum:\*  
24: em\_gss\_hum:\*  
25: em\_gss\_hum:\*  
26: em\_gss\_hum:\*  
27: em\_gss\_hum:\*  
28: gb\_gss1.\*

29: gb\_gss2.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4528.5	96.4	3143	11	AK016546	AK016546 Mus muscu
2	4403	93.7	5517	11	AK029788	AK029788 Mus muscu
3	2833	60.3	3327	11	AK050021	AK050021 Mus muscu
4	2823	60.1	3457	11	AK078301	AK078301 Mus muscu
5	2820	60.0	3376	11	AK050023	AK050023 Mus muscu
6	1693.5	36.0	1042	12	BM557438	BM557438 AGENCOURT
7	1542	32.8	957	13	BO675006	BO675006 AGENCOURT
8	1539.5	32.8	910	13	BO675260	BO675260 AGENCOURT
9	1505.5	32.0	1041	12	BI084090	BI084090 602869453
10	1430	30.4	968	13	BO671635	BO671635 AGENCOURT
11	1427	30.4	910	13	BO671276	BO671276 AGENCOURT
12	1397	29.7	919	13	BM390898	BM390898 BX390898
13	1370.5	29.2	853	12	BI223892	BI223892 602941035
14	1354	28.8	1082	13	BU239476	BU239476 603322338
15	1343	28.6	753	9	AL040398	AL040398 DKFZ64348
16	1342	28.6	789	10	BM709118	BM709118 602675382
17	1342	28.6	3726	11	AK039652	AK039652 Mus muscu
18	1325	28.2	1090	12	BM910838	BM910838 AGENCOURT
19	1286	27.4	746	14	BY751026	BY751026 BY751026
20	1277.5	27.2	985	13	BO068650	BO068650 AGENCOURT
21	1276	27.1	831	12	BM974324	BM974324 602844075
22	1260	26.8	802	14	CD354396	CD354396 UI-M-GMO-
23	1248	26.6	735	13	BU447277	BU447277 60376685
24	1242.5	26.4	788	10	BM163397	BM163397 602338360
25	1242.5	26.4	1021	12	BM974587	BM974587 602844820
26	1225	26.1	741	14	CB595102	CB595102 AGENCOURT
27	1216	25.9	855	12	BI084885	BI084885 60269453
28	1210.5	25.8	993	10	BM259714	BM259714 602380072
29	1201	25.6	762	14	CD352521	CD352521 UI-M-GIO-
30	1196	25.4	932	14	CA790315	CA790315 AGENCOURT
31	1192	25.4	695	10	BM715866	BM715866 602675282
32	1171	24.9	910	10	BM479035	BM479035 602526111
33	1170	24.9	897	13	BM784154	BM784154 602108015
34	1136	24.2	897	13	BU191638	BU191638 AGENCOURT
35	1130.5	24.1	674	13	BU229303	BU229303 603800110
36	1127.5	24.0	774	14	CB233750	CB233750 AGENCOURT
37	1127	23.8	669	10	BM390103	BM390103 602415944
38	1120	23.8	1002	10	BM733691	BM733691 601569288
39	1117	23.8	674	10	BM888665	BM888665 601513061
40	1099	23.4	639	10	BM165205	BM165205 602444074
41	1081	23.0	658	10	BM388695	BM388695 601286345
42	1080.5	23.0	853	10	BM694553	BM694553 602081767
43	1073	22.8	661	10	BM799580	BM799580 601592450
44	1066	22.7	617	10	BM789604	BM789604 601481552
45	1066	22.7	845	13	BU216915	BU216915 603109373

## ALIGNMENTS

RESULT 1  
AK016546  
LOCUS  
DEFINITION  
AK016546 3143 bp mRNA linear HTC 05-DEC-2002  
Mus musculus adult male testis cDNA, RIKEN full-length enriched  
library, clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog  
[Homo sapiens], full insert sequence.  
ACCESSION  
AK016546  
VERSION  
AK016546.1 GI:12855334  
KEYWORDS  
HTC, CAP trapper.  
SOURCE  
Mus musculus  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	TITLE	JOURNAL	PUBLISHED	AUTHORS
1	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)	JOURNAL	99279253	
2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)	JOURNAL	20499374	
3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer Genome Res. 10 (11), 1757-1771 (2000)	JOURNAL	20530913	
4	Kawai, Y., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Goto, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Cassavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabili, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barish, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., But, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotamann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P., Rao, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Williams, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)	JOURNAL	21085660	
5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)	JOURNAL	21127851	
6	(bases 1 to 3143)	JOURNAL		
7	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Ota, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (10-JUN-2000) Yoshinori Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome	JOURNAL		

COMMENT

Exploratory Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel.:81-45-503-9222,  
Fax:81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGAGAAGGAAGATTCACCAAGCCTCTTTTCTTTTCTTNN 3'] cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. Second  
strand cDNA was prepared with the primer adapter of sequence [5'  
GGAGAGAGATTCTCGAGTTAATTAAATTATCCCCCCCCC 3']. cDNA was cleaved  
with BamHI and XhoI. cDNA of size comprised longer than 7 kb was  
selected before cloning. Vector: a modified plasmidscript KS(+/-) after  
bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3'  
end: BamHI. Host: DH10B.

FEATURES

SOURCE

Location/Qualifiers

1..3143

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:4932434F09"

/db\_xref="MGI:1907554"

/db\_xref="taxon:10090"

/clone="4932434F09"

/sex="male"

/tissue-type="testis"

/clone\_id="Riken full-length enriched mouse cDNA library"

/dev\_stage="adult"

327...3005

/note=unlabeled protein product: DIPPTIDYL PEPTIDASE 8  
homolog (Homo sapiens) (SPRR109HM5, evidence: FASTY,  
95.%ID, 100%length, match=2649)

putative"

/codon\_start=1

/protein\_id="BAB30295.2"

/db\_xref="MGI:26382128"

/db\_xref="MGI:1921638"

/translation="MKIPSGRCMAAMAELEQUCVEIFETACEEGNGESODPKLER  
FYVERYSWSQLKDLADTRKYHGYYAKAHADPMFAVARTDPDRPHSDRVYLLMSGEN  
RENTLIFYSELPKPINNAVLMLSMKPLDIEFOATLDGYMSREBELREKRIGTVGI  
AAVDYHSGSTPLFOAGSGIVYIKDGPHGFOTOPRLNVTNSCPNIMDKPCPADGI  
PDWIAFHSHNDIWISNLVTEERIRITYVHNELNMEDPDSAGVATPVLOEPFRGSY  
VMXCOAPERTPSGGKILRIIXENDSEEVIIHTSPMLTPRADASRYKTKANGK  
VTRKSEIIVDAAGGIIDVIDKELVQFEILFECEVIETIAGACTPEGRKAWSLTLDSS  
QTUHQVLISELPFIPEVDAMDORQLIESVPSTVPLIIYEETDIWINIHDFHFV  
POHDEIDEIFASECTEDGPRHLVKITSLIKESYKKSOGGLAPSPFCPIKEIITI  
TSGEWELHRGHGINIWDPEARKLVEFGTOSPLEHLVYTSAANPEVRLTDRGYS  
HSCSLASHCDPLFSKYSNQNPCHCVSLKCSPEDDVHKTFEMATILDSAGLPGDY  
TLPIEFSEFTGTFTLYGMLYKPHDLOPKKKYPTVLFYGGPOVALNNPFKVXYR  
LMTLASIGYVVVVINDRSGCRBLKEGAKCMGOGLTIIDPOUEGLOLAISOYFDIL  
DRGICGMSYCYGLSLMALMORSDIRRVATAGAPVTMIITDYGTERTYTGHPDNOE  
GYLVGSVAQMAEKPPESPKNRLILLHGLDENVPARHSILTSLVRAGRPYDIQIYPO  
ERRHSIRPESGEHYEHLTHLYLOENNGSRALAAKVTI"

BASIC COUNT

BASE COUNT

894 a

688 c

773 g

768 t

ALIGNMENT SCORES:

Score: No.:

0

Length:

3143

Percent Similarity:

4528.50

Matches:

845

Best Local Similarity:

97.73%

Conservative:

18

Query Match:

95.70%

Mismatches:

19

DB:

96.35%

Indels:

1

Gaps:

1

US-10-070-464-1 (1-882) x AK016546 (1-3143)

1 Metc1A



[illegible]

REFERENCE	1	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Komno, H., Akechi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Cavaletto, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikidol, I., Pesole, G., Quackenbush, J., Schiraldi, L. M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Bairdrell, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Monbaerts, P., Notodone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Saeki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyono-Oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wyšnawski-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohetsuki, S. and Hayashizaki, Y.
TITLE		Functional annotation of a full-length mouse cDNA collection
JOURNAL		Nature 409 (6821), 685-690 (2001)
MEDLINE		21085660
PubMed		11217851
REFERENCE	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
TITLE		(bases 1 to 5517)
JOURNAL		Aachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imtani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Onno, M., Osato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE		Direct Submission
JOURNAL		Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers
FEATURES		1. 5517
source		/organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:4930560C15" /db_xref="taxon:10090" /clone="4930560C15" /sex="male" /tissue_type="testis" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 2221. 4837 /note="DIPEDIPDYI, PEPTIDASE 8 homolog [Homo sapiens]"
misc_feature		

(SPT|Q9HWS, evidence: FASTY, 95.7%ID, 100%length, match=2649)  
 putative"

BASE COUNT 1514 a 1224 c 1230 g 1549 t

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	5517
Score:	4403.00	Matches:	827
Percent Similarity:	95.48%	Conservative:	17
Best Local Similarity:	93.55%	Mismatches:	17
Query Match:	93.68%	Indels:	23
	11	Gaps:	2

US-10-070-464-1 (1-882) x AK029788 (1-5517)

```

Qy 1 MetAlaAlaMetGluThrGluGluLeuGluValGluIlePheGluThrAlaAspCys 20
Db 2248 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTCAGATATTGAAACTGCAGAGTGT 2307
Qy 21 GluGlu--AsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArg 39
Db 2308 GAGGAGGAAATGCGATCATCAGAGCCGACCCAACTGAGACCTTTTATGTGCAACGG 2367
Qy 40 TyrSerTrpSerGlnLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMet 59
Db 2368 TATTCTGGAGTCAAGCTGAAAGAGCTGCTGTGATACAGAAATACACAGGCTACATG 2427
Qy 60 MetAlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHis 79
Db 2428 ATGGCTAAGGCGCCACATGACTTATGTTTGAAGAGAGCGATCCAGATGGGCTTAC 2487
Qy 80 SerAspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgLysGluAsnThrLeuPheTyr 99
Db 2488 TCAGACAGGCGTATATTACCTTGCCCATGTCTGTGTGGAACAGAGAAATACATATTAT 2547
Qy 100 SerGluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeu 119
Db 2548 TCTGAATCCCTTAAACCATCAACAGACAGACAGCTTAAATGCTTCTCGAAGCCCTT 2607
Qy 120 LeuAspLeuPheGlnAlaThrLeuAspArgLysMetTyrSerArgGluGluLeuLeu 139
Db 2608 TTGGATCTTTTTCAGGCGACACTAGACTAGTGGATGTTCTCAGAGAGAGACTATCTC 2667
Qy 140 ArgGluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySer 159
Db 2668 AGAAGAAAGAAAGCCCATTTGGAAACCGTGGGATCCGAGCTTATGATTACACCCAGAAAC 2727
Qy 160 GlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGln 179
Db 2728 GGAACATTCCTGTTCAAGCTGGCAGCGGCAATTACATCAAAAGATGAGAGACCAT 2787
Qy 180 GlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArg 199
Db 2788 GGATTTACACACAGACTTGGCGGCCCAATTTAGGAACTAGTGTGCCCAAAATACGA 2847
Qy 200 MetAspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAsp 219
Db 2848 ATGGATCCAAATATATGCGCCGTGATCCAGACTCGATAGCTTTCATTCACAGCAATGAT 2907
Qy 220 IleTrpIleSerAsnIleValThrArgGluGluArgLysLeuThrTyrValHisAsnGlu 239
Db 2908 ATTGGATATCAAACTCGTAAACAGGAGAGAGCGAGATCATCATGCTACATAGAG 2967
Qy 240 LeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 259
Db 2968 CTAGCCAACTAGAGAGAGATCCAGATCAGCTGTGTGCGACCTTTGTTCAGAA 3027
Qy 260 GluPheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGly 279
Db 3028 GAATTTGACAGATATCTGTGGTACTGTGGTGTGCCCAAGGAGAAAGAACTCCCTAGTGT 3087
Qy 280 GlyLysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHis 299

```

```

Db 3088 GGTAAATTTCTAGATTTCTCATGACAAATAATGATGATCGAGTGGAGATTATTCAT 3147
Qy 300 ValThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyr-ProLysThrGly 319
Db 3148 GTTAGCTCCCCCATGTTGGAAACAGAGGCGAGATTCCTTTCGTTATCCCAAAA----- 3202
Qy 319 YThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArg 339
Db 3203 -----CA 3204
Qy 339 GluIleLeuAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyVal 359
Db 3205 GATTATAGATGTCATAGATTAAGAACTGTTCAACTTCGAGATTCGTTTGAAGGAGT 3264
Qy 359 IGluTyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaTyrPheIleLeuLe 379
Db 3265 TGAATATATTTCAGAGCCGAGTGAATCCAGAGGAGAAACATGCTGCTCATCTACT 3324
Qy 379 uAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProVal 399
Db 3325 AGACCGTTCACAGACTCACTGCAGATGTTCTGATCTCCCTGAGTTATTCATCCAGT 3384
Qy 399 IGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrPr 419
Db 3385 AGAAGATGATGCCATGACAGACAGACAGACTTATAGACTTCTGACTCTGTGACACC 3444
Qy 419 GluIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisVal 439
Db 3445 ACTGATCATCTATAGAAACAAACAGACATCGATGAATTAATTCACAGATATTTTCATGT 3504
Qy 439 LProGlnSerHisGluGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGly 459
Db 3505 TTTTCTCCAACTAGAGAGAAATGAGTTATGATTTATTTGCTGTGATCAAAACAGG 3564
Qy 459 YPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerTyrLysArgSerSe 479
Db 3565 TTTTGGTATCTGATATAAATACATCATCTTTTAAAGAGAGCAATTAACGCTCAG 3624
Qy 479 rGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThr 499
Db 3625 TGTGGACTACTCTCCCAAGTAGTTCAAGTGTCTTATCAAGAAAGAAATTAACATTAC 3684
Qy 499 rSerGlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArg 519
Db 3685 CAGTGTGAATGGAGAGTACTGGCGGACATGATCTTAATATCTGGGTGATGAAGCCAG 3744
Qy 519 gArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValIle 539
Db 3745 AAAGCTGTGATCTTTGAAGCACCAAGACTCTCTTTGGAGCATCACTGTACGTGAC 3804
Qy 539 LserTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCys 559
Db 3805 CAGTTATGCAAAACCTGAGAGAGTGGAGCTGACTGACGCTGCTACTACACTCCTG 3864
Qy 559 sCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHis 579
Db 3865 CTGCTTACCGCGGATTTGATCTTTCATTAAGTAAGTAAACAGCAACAGAAATTCACA 3924
Qy 579 sCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGly 599
Db 3925 CTGTGTGCTCCTTCAAACTCTCAACTCTGAGAGATGACCCAGTTTCATTAACAAAGGA 3984
Qy 599 uPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIle 619
Db 3985 ATTTGGGCGACCAATTTGGATTACAGAGGCTCTCTTCGACTACACCCCTCAGAAAT 4044
Qy 619 ePheSerPheGluSerThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAs 639
Db 4045 TTTTCTTTTGAAGATACATGATTTTACATGATGAATGTGATTAAGCTCTATGA 4104
Qy 639 pLeuGlnProGlyLysLeuTyrProThrValLeuPheIleTyrGlyGlyProGlnValGly 659

```

Db	4105	CCTAACACCTGGAAAGAAATACCCACACCTGTATTATCATATATAGTGGTCCCAAGTCA	4164
Qy	659	mLeuVal1asnAsnArgPheLySGlyVal1yStryPheArgLeuMenth1Leu1aSerLe	679
Db	4165	GCTGGTGAACATCGGTTTAAAGGAGTCAGAGTATTTCCGCCGAACACCTGGCCCTCC	4224
Qy	679	uG1yTyrrVal1Val1Val11leAspAsnArgQ1ySerCysH1sarG1yLeuLyPheG1	699
Db	4225	GGGTTATGTGTTGGTGGATGATGACACAGGGAGTCCCTGACCGAGGACTTAAATTGA	4284
Qy	699	uG1yVal1aPheLyStryLySMeG1yGln1leGln1leAspAspG1Val1G1uG1yLeuG1	719
Db	4285	AGGCCCTTTAAATATTAATGATGGCTCAATAGAAATCATGATCATCAAGTGAAGACTCA	4344
Qy	719	nTyrrLeuVal1aSerArgTyrrAspPhe11aAspLeuAspArgVal1G1y1leH1SG1yTrnS	739
Db	4345	GTACTTCAGATTCCTGATGATCTTATGACTTGATGAGGGGCAATCCACGGCTGGTC	4404
Qy	739	rTyrrG1yG1yTyrrLeuSerLeuMetAla1euMetGlnArgSerAsp11lePheArgVal1	759
Db	4405	CTATGTGGCTACCTCTCCCTGATGGCATTAATGACAGAGTCCGATATCTTCCGGGTGC	4464
Qy	759	a11eAlaG1yAlaProVal1ThrLeuTrp11lePheTyrrAspThrG1yTyrrGlnArgTy	779
Db	4465	TATTCCTGGGGCCCAAGTCAACCTGTGATCTTCTATATATACAGGATACAGGAGCTCA	4524
Qy	779	rMetG1yH1sProAspG1naSnG1uG1nG1yTyrrTyrrLeuG1ySerVal1AlaMetGlnAl	799
Db	4525	TATGGGTCACTTCACCAAGATGAACAGGGCTACTACTTAGGATTTGGCCATTCAGC	4584
Qy	799	aG1uLySpheProSerG1uProAsnArgLeuLeuLeuH1SG1yPheLeuAspG1uAs	819
Db	4585	GGAGAGATTCCTCCCTCAGAACCAACCGGTTACTCTTGTCATGGGTTCTTGATGAAGA	4644
Qy	819	nValH1sPheAlaH1sTrnSer11leLeuLeuSerPheLeuVal1ArgAlaG1yLySProTy	839
Db	4645	TGTTCACTTGACACACCAAGTATATTTGCTGAGTTTTTAGTAGGGCTGGAAACCCTA	4704
Qy	839	rAspLeuGln11eTyrrProGlnG1uArgH1sSer11leArgVal1ProG1uSerG1yGlnH1	859
Db	4705	TGACTTACGATCTATCTCCAGAGAGGACGACATCAGAGTTCCTGAGTCTGGAGACA	4764
Qy	859	sTyrrGlnLeuH1sLeuLeuH1sTyrrLeuGlnGlnuSnLeuG1ySerArg11leAla1le	879
Db	4765	CTATTAACCTGACCTGCTCCACTCACTTCCAGAGAACCTTGATGCGTATCGTCTCT	4824
Qy	879	uLyVal11le 882	
Db	4825	GAAATGATA 4834	
RESULT 3			
AK050021		3327 bp mRNA linear HTC 05-DEC-2002	
LOCUS			
DEFINITION		Mus musculus adult male liver tumor cDNA, RIKEN full-length	
ACCESSION		enriched library, clone: C73003D12 product: DIPPTIDYL PEPTIDASE 9	
VERSION		AK050021	
KEYWORDS		AK050021.1 GI:26340743	
SOURCE		HTC; CAP trapper.	
ORGANISM		Mus musculus (house mouse)	
REFERENCE			
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
JOURNAL		Carninci, P., and Hayashizaki, Y.	
MEDLINE		High-efficiency full-length cDNA cloning	
PUBMED		Meth. Enzymol. 303, 19-44 (1999)	
REFERENCE		99279253	
AUTHORS		10349636	
TITLE			
		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
		Itou, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
		Normalization and subtraction of cap-trapper-selected cDNAs to	

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Katsuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawat, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Komoto, H., Adachi, J., Fukuda, S., Arakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischnann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nishikido, I., Peele, G., Quackenbush, J., Schriml, L.M., Stabli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Camilleri, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotamner, C., Hume, D.A., Kamlay, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Togo, Oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilmink, L., Wyshew-Boris, A., Yoshida, K., Haegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	6 (bases 1 to 3327) Adachi, J., Aizawa, K., Akinura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirokawa, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Submitted (16-JUL-2001) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-research@riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.



Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>

URL: <http://fantom.gsc.riken.go.jp/>

Location/Qualifiers

1. 3327

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:C730003D12"

/db\_xref="taxon:10090"

/clone="C730003D12"

/sex="male"

/cell\_type="tumor"

/tissue\_type="liver"

/clone\_1ib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

176..2764

/note="unnamed protein product; DIPPTIDYL PEPTIDASE 9

homolog (Homo sapiens) (SETR|AAU47179, evidence: FASTY,

92%ID, 99.6%length, match=2580)

putative"

/codon\_start=1

/protein\_id="BAC34034.1"

/db\_xref="GI:26340744"

/translation="MCSGVSPVEOVAAGMDTAAFCVQKHSMDGLNSIHGRKSS

GLIVSKAPDFOVQKPDSPGSHRLYLGMPSRENSLLYSEI PKVVRKALLL

SKOMLDHFQATPHGVSEBELERKRLGVGITSDFSEGLFLFQSNLFLH

CDGKGKMGVPMKPLEIKTCCSGRPMKICPADPAFSPINSDLVANIERGE

RUTFCGHSAGVLDNPKAGVATFVIOEFPDRCPAMFSCPTASNGSGGLTLILY

BEVDSFVHTVPSPALERKTSRYRRTGSKPKIALKALIEOTHQGRTVSCC

KELVDPFSLFPKVEYIARAGTRDCKYAMAFILDRPOQLVLPLAFIPAVESE

AQRQAARVPRKVPFVLYEVTNWIMVHDI FHPPEAQEQDFCLRAECCTGF

CHLYRVTELKTKDYTEPLSPTEDEPCPIKEEVALTSGEMVLSRGSXIWNEQ

TLVYVFOGKTQPLEHNLVYVESAGEIVRLTTLGFSHSCSDFMVFYSVS

TPPCVHYKLSGDDPLKORFPMASMEANCPDVPYPERI FPHHTRADVLGMI

YKPHLQGRKHPVLYFGYKPOVLYUNNSFKGIKYLRLNTLASGYAVVYDGGSC

ORGHFEKALKQKQKQVEIEDYVEGLQYIAEKYFIDLSRAVHKGSTGGFLSLGLI

HKPVFKVAILAGAPVTVMAYDTGTERYMDVPENNOQSEKGSVALHEKLPENPR

LILHGFLENVHFPHYFNLVSQLIRAGKPYQLQIYPMERHSIRCESEHVEYVTLH

FLOEHL"

BASE COUNT 702 a 1015 c 919 g 691 t

ORIGIN

Alignment Scores:

Pred. No.: 1.45e-304 Length: 3327

Score: 2833.00 Matches: 511

Percent Similarity: 76.79% Conservative: 134

Best Local Similarity: 60.83% Mismatches: 193

Query Match: 60.28% Indels: 2

DB: 11 Gaps: 2

US-10-070-464-1 (1-882) x AK050021 (1-3327)

QY 35 PheTYrValIGluArgTYSerThrpSerGlnLeuLYsLYsLeuAlaAspThrArgLYs 54

Db 242 TTTCTGTGGCAAGACTCGTGGAGTGCGCGAGCATTAATCCAGCGCTGCAAG 301

QY 55 TyrHisGlyTYrMetMetAlaLYsAlaProHisAspPheMetPheValLYsArgAsnAsp 74

Db 302 TCCCTGGGGCTCATTTGTGACAAAGGCCCCCAAGACTTTCAGTTTGTGACAAAGCTGAC 361

QY 75 ProAspArgLYsProHisSerAspArgLYsTYrTYrLeuAlaMetSerGlyGluAsnArgGlu 94

Db 362 GAGTCTGGCCCCCACTTCACCGCTCTATTAACCTCGGATATCCTTACCGCAGCGGTGAG 421

QY 95 AsnThrLeuPheTYrSerGlnLeuLeuProLYsThrIleAsnArgAlaAlaValLeuMetLeu 114

Db 422 AACTCTCTCTACTCCGAGATCCCAAGAAAGTGCGGAGAGAGCGCCTGCTGCTG 481

QY 115 SerTrpLYsPProLeuLeuAspLeuPheGlnAlaThrLeuAspTYrGlyMetLYsSerArg 134

Db 482 TCTCGAAGACAGATGTGACCACTTCCAGGCCACACCCACCATGTGTCTACTCCCA 541

QY 135 GluGluGluLeuLeuArgLYsArgLYsArgLYsArgLYsArgLYsArgLYsArgLYs 154

Db 542 GAGGAGGAGACTCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 601

QY 155 TyrHisGlnGlySerGlyTYrPheLeuPheGlnAlaGlySerGlyLYsTYrHisValLYs 174

Db 602 TTCACAGTGAGAGCGGCTCTTCTCTTCCAGGCCAGCAAAAGCTGTTCACCTGACAG 661

QY 175 AspGlyGlyProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194

Db 682 GATGTGGCAAGATGTGCTTATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721

QY 195 CyProAsnIleArgMetAspProLYsLeuCyProAlaAspProAspTrpIleAlaPhe 214

Db 722 TGTCTGGGCGACGATGAGCCCAAAATCTGCGGAGACCTGCTGCTGCTGCTGCTGCTGCT 781

QY 215 IleHisSerAsnAspIleTrpIleSerAsnIleValThrArgGluGluArgArgLeuThr 234

Db 782 ATCAACAAACAGATCTGTGGTGGCAACATCGAGACTGGGAGAGAGAGAGAGAGAGAGAG 841

QY 235 TyrValHisAsnGluLeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThr 254

Db 842 TTTCTGTACCAAGGTTACAGCTGTGTCTCTGCAACATCCAAATTCACAGCGGTGACAC 901

QY 255 PheValLeuGlnGluGluPheAspArgTYrSerGlyTYrTrpTrpCyAspProLYsAlaGlu 274

Db 902 TTTGTATCATCAAGAGAGAGTTCACAGCGCTTCACTGGGTGTGTGTGTGTGTGTGTGTGT 961

QY 275 ThrThrProSerGlyGlyLYsLYsLeuAlaGlyLeuTYrGluGluAsnAspGlySer 293

Db 962 TCGAAGGCTCGAAGGCTCAAGACGCTCCATCTATAGAGAACTGACGAGTGT 1021

QY 294 GluValIGluIleIleHisValThrSerProMetLeuGluGluThrArgArgAlaAspSerPhe 313

Db 1022 GAAAGTGAGATATTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081

QY 314 ArgTYrProLYsThrGlyThrAlaAsnProLYsValThrPheLYsMetSerGlyIleMet 333

Db 1082 CCTACCCCAAGACAGGACGAGCAAGAACCCCAAGATTGCCCTGAGCTGAGCTGACG 1141

QY 334 IleAspAlaGluGlyArgIleIleAspValIleAspLYsGluLYsIleGlnProPheGlu 353

Db 1142 ACGACCATCAAGGCAAAATCTGTCAAGCTCGAGAGAACTGTATCAGCATTCACAC 1201

QY 354 IleLeuPheGluGlyValGluTYrIleAlaArgAlaGlyTYrTrpThrProGluGlyLYsTYr 373

Db 1202 TCCCTTTTCCCAAGTGTGAGTACATCGCCGAGGCTGCTGAGACACGAGGACCAATAT 1261

QY 394 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393

Db 1262 GCTGAGCCATGTTCTGTGACCGCTGCCAGCAACGCTTACCTGTCTCTGCTGCTGCTGCT 1321

QY 394 GluLeuPheIleProValIGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413

Db 1322 GCTCTTCTATCTCCGCGGTGTGAGAGTGAAGGCCACGCGGACGACCTGCCAGACCGCT 1381

QY 414 ProAspSerValThrProLeuIleIleTYrGluGluGluThrThrAspIleTrpIleAsnIle 433

Db 1382 CCCAAGATGTGACAGCTCTTGTCTATCTATGAGAGATCAACAAATGTCTGATCAACGTC 1441

QY 434 HisAspIlePheHisValPheProGlnSerHisLYsGluGluIleGluPheIlePhe 452

Db 1442 CACGACATCTTCCACCGCTTCTGACGCTGAGGCGACGAGCACTTTTGTGTTCTGCTGT 1501

QY 453 AlaSerGluCyLYsThrGlyPheArgHisLeuTYrLYsIleThrSerIleLeuLYsGlu 472

Db 1502 GCCAAGATGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1561

QY 473 SerLYsTYrLYsArgSerSerGlyGlyLeuProAlaProSerAspPheLYsCyAspProIle 492

Db	1562	AAGAGCTATGACTGGACGGAAACCCCTCAACCCCTTACAGAAAGATGAATTAAATGCCCCATC	1621
Qy	493	LysGIuGIuIIeAlaIIeThrsErGIyGutPrGIuValIleuGIyArgHIgIySerAn	512
Db	1622	AAGAGGAGAGTGCCTGACCTGACCAAGTGGGAGGAGGCTTGTGAGGCAATGAGCTCCAAAG	1681
Qy	513	IIeGIuValAspGIuValArgArgIleuValTYrPheGIuGIyThThIyAspSerProIeu	532
Db	1662	ATCTGGGGTCAACGACACGACGAAAGCTGGGGTACTTTCAAGGACAAAGACACACCGCTG	1741
Qy	533	GIuHIshIeLeuTYrValIeSerTYrValAsnProGIyGIuValThrArgIleuThrAsp	552
Db	1742	GAACATCAACCTCTATGTGGTGAGCTACAGCATGACGACGAGGATGTGGGGCTCACAG	1801
Qy	553	ArgGIyTYrSerHIssErCyseCysIIesErGIuHIscYsAspPhePheIIeSerIyTYr	572
Db	1802	CTCGGCTTCTCCACAGCTGCTCCATGAGCCAGACGCTTGCATGTTGTAGATCAACAC	1861
Qy	573	SerAsnGIuIyAsnProHIscYsValSerIyTYrIyIeSerSerProGIuAspAsp	592
Db	1862	AGCACTGTGAGACACCCACCCCTGTGATGATGTCMAAGCTGAGCGGCCCCGATGATAC	1921
Qy	593	ProthTYrCysIyThIyLeuGIuPheTrpAlaThrIIeIleAspSerAlaGIyProIeuPro	612
Db	1922	CCACTGCACAAAGCAACCAACGCTTCTGGGCGACAGATGAGAGGCGCAATTTGCCCCCA	1981
Qy	613	AspTYrThrProProGIuIIePheSerPheGIuSerThrThGIyPheThIeIyTYrGIy	632
Db	1982	GACTATGTGCCCCCTGGATCTTCCACTTCCACACCCGTCGACAGCTGACGCTTACGGC	2041
Qy	633	MetLeuTYrIyPheProHIAspIleuInProGIyIyIyIyTYrProThValIleuPheIIe	652
Db	2042	ATGATCTCAAGCCACACACCCCTGCACTGGGAGAGAGCAACCCACGTGCTCTTGTGC	2101
Qy	653	TYrGIyGIyProGIuValGIuIleuValAsnAsnArgPheIyValIyTYrPheArg	672
Db	2102	TATGGGGGCCCAAGAGTGCAGTGGTGAAACACTCTTAAAGGATCAAAATACCTGGGG	2161
Qy	673	LeuAsnThrLeuAlaSerIleuGIyTYrValValValIIeAspAsnArgGIySerCys	692
Db	2162	CTAAATACACTGGCAATCTCTGGGCTATGCTGTGGTGATGATGATGTGCGGGCTCGGT	2221
Qy	693	HisArgGIyLeuIyPheGIuGIuValaPheIyTYrIyIyMetGIyGHIleGIuIIeAsp	712
Db	2222	CAGCGGGGCTGCACTTGCAGGGGGGCCCTGAAATAATCAAAATGGCGAGGTGAGATTGAG	2281
Qy	713	AspGIuValGIuGIyLeuGIuInTYrIleuAlaSerArgTYrAspPheIIeAspIleuAspArg	732
Db	2282	GACCAAGTGGAAGCGCTTGACAGTACGTGGCTGAGAGTATGGCTTACATTGACTGAGCCGA	2341
Qy	733	ValGIyIIeHIsgIyTrpSerTYrGIyTYrIleuSerIleuMetAlaIleuMetGIuArg	752
Db	2342	GTCGGCATTCATGGGTGGTCTTCAACGGCGGCTTCCCTCACTCATGAGGGCTCATCACAAAG	2401
Qy	753	SerAspIIePheArgValAlaIIeIleIleIleGIyAlaProValThrIleuTrpIIePheTYrAsp	772
Db	2402	CCACAAGCTTCAAGGTAGCCATTGGGGCGCTCTGTGCATCTGTGTGATGGCTTATGAC	2461
Qy	773	ThrGIyTYrThrGIuArgTYrMetGIyHIspProAspGIuAsnGIuGIuInGIyTYrTYrIleu	792
Db	2462	ACAGGGTACACGGAACGATACATGATGTCCCGGAAATAACACAGCAAGGCTATAGGCA	2521
Qy	793	GIyIerValAlaIleMetGIuIleGIuIyPheProSerGIuProAsnArgIleuIleuIleu	812
Db	2522	GGGTCTTGAGCCCTGCATGTGAGAGAGCTGCCCAATGAGCCTAACCGGCTCTTATCTCTC	2581
Qy	813	HisGIyPheIleuAspGIuAsnValHIspPheAlaHIsthTrsIIeIleuIleuSerPheIleu	832
Db	2582	CACGGCTTCTGTGAGAGAAAGTTCACTTCTTCCACAAATTTCTGTGTGCCAGCTG	2641
Qy	833	ValArgAlaGIyIyPheProTYrAspIleuInIleTYrProGIuGIuArgHIssErIIeArg	852

Ox	853	ValProgluterglyglyluhiSTyrgluLeuhisIeuleuhiSTyrlEuGIInGluAbnleu	872
D6	2702	TGCCCGCAGATCCGAGAGACATTACGAGGTGACGCTGTCGACTTTTCTGCAGGAACACCTG	2761
RESULT 4			
LOCUS	AKO78301	3457 bp	mRNA linear HTC 05-DEC-2002
DEFINITION	Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430584g11 product:DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens), full insert sequence.		
ACCESSION	AKO78301		
VERSION	AKO78301.1	GI:26347124	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus Eumetazoa; Chordata; Craniata; Vertebrata; Eutelestomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636		
AUTHORS	Garcini, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159		
JOURNAL MEDLINE PUBMED	Shibata, K., Itoh, M., Aizawa, K., Nagaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matchiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawaj, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861		
JOURNAL MEDLINE PUBMED	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Salto, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Salto, R., Kadote, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Knehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Perozo, G., Quackenbush, J., Schriml, L.M., Straubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, O., Boffelli, D., Bojunga, A.N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Ball, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mommaerts, P., Nordone, P., Rung, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Saeki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Scorch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohzuki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851		
JOURNAL MEDLINE PUBMED	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
REFERENCE			



Db 1208 TCCCTTTTCCCAAGGAGTACATGCGCCGCTGCGACAGCGGCAATAT 1267  
 Qy 374 AATATPSeTLeuLeuAAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393  
 Db 1268 GCTGGGCGCATGTTCTGTATCGCTCCCGCAAGAGGCTTACGTTCTCTCAACCCCT 1327  
 Qy 394 GluLeuPheIleProValGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413  
 Db 1328 GCTCTTTCATCCCGCGCTGAGAGTGAAGCCAGCGGCGACAGCTGCGCAAGCCGCT 1387  
 Qy 414 ProAspSerValThrProLeuIleIleTyrGluGluThrThrAspIleThrIleAsnIle 433  
 Db 1388 CCCAAGAAATGAGCCCTTGTATCATCTATGAAGAGTCAACATGCTGATGATCAACGTC 1447  
 Qy 434 HisAspIlePheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePhe 452  
 Db 1448 CACGACATCTTCCACCCGTTTCTCAAGGCTGAGGCGCAGACAGACTTTGTTCTTCCTG 1507  
 Qy 453 AlaSerGluCysAlaYsThrGlyPheArgHisLeuTyrIleTyrSerIleLeuValGlu 472  
 Db 1508 GCCAAAGAAATGAAAGATGGCTTCTGACACTGTACAGGCTCACAGTGAACCTTAAAC 1567  
 Qy 473 SerTyrTyrIleArgSerSerGlyGlyLeuProAlaProSerAspPheIleYsCysProIle 492  
 Db 1568 AAGGACTATGACTGAGACGGAACCCCTCAGCTTACGAAGATGAGTTTAAGTCCCATC 1627  
 Qy 493 LysGluGluIleAlaIleThrSerGlyGluTyrGluValLeuGluValArgHisGlySerAsn 512  
 Db 1628 AAGGAGAGAGTGCTGCTGACCACTGAGAGTGGAGAGTCTTTCGAGGCGATGCTCCAA 1687  
 Qy 513 IleGlnValAspGluValArgArgLeuValTyrPheGluGlyThrIleAspSerProLeu 532  
 Db 1688 ATCTGGCTTACCAAGAGCAAGAGAGCTGGTGTACTTTCAGAGTCAAGAGACACACCGCTG 1747  
 Qy 533 GluHisIleLeuTyrValIleSerTyrValAsnProGlyGluValThrArgLeuThrAsp 552  
 Db 1748 GAACATCACCCTCATATGGTGTAGCTACAGTCAAGGAGGAGATCTGCGCTCACCAAG 1807  
 Qy 553 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPheIleSerIleTyr 572  
 Db 1808 CTCGGCTTCCCAAGCTGCTCCATGAGCCAGACAGCTTGACATGTTCTGAGTCACTAC 1867  
 Qy 573 SerAsnGlnLysAsnProHisCysValSerLeuTyrLeuLeuSerSerProGluAspAsp 592  
 Db 1868 AGCAGGTGAGACACGCTGCTGTACATGTGTACAGTGAAGCGGCCCCCATGTATGAC 1927  
 Qy 593 ProThrCysIleThrIleGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuPro 612  
 Db 1928 CCACGTCAACAAGCAACACGCTTCTGGCGCAGCATGATGAGGACCAATTGCCCCCA 1987  
 Qy 613 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 632  
 Db 1988 GACTATGTGCCCCCTGAGATCTTCCACTTCCACACCCGTCAGACGTCGACGCTCTACG 2047  
 Qy 633 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 652  
 Db 2048 ATGATCTTACAGCCACACACCTGCAACTTGGAGAGAGACCCCACTGCTCTTTCCTG 2107  
 Qy 653 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValIleTyrPheArg 672  
 Db 2108 TATGGGGGCCACAGAGTGCAGTTGTGGAACAATCTCTTAAGGGCATCAAAATACCTGG 2167  
 Qy 673 LeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCys 692  
 Db 2168 CTAAATATACCTGGACCTTGGCTATGTCTGTGTGTATGATGATGTGTGGGGCTCTCT 2227  
 Qy 693 HisArgGlyLeuLeuPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluLeuAsp 712  
 Db 2228 CAGCGGGGCTGCATCTGAGGGGGCGCTGMAAAATATGAGCCAGGTGAGATTGAG 2287  
 Qy 713 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 732

Db 2288 GACCAAGTGAAAGCTTGCAATACGTGCTGAGAAATATGGCTTACTTACCTTGACCCCA 2347  
 Qy 733 ValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752  
 Db 2348 GTCCGACATCATAGGCTGTGCTTACAGCGGCTTCTCTCATCTATGAGGGCTCATCCAAAG 2407  
 Qy 753 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAsp 772  
 Db 2408 CCACAAGTGTTCAGAGTACCATTTGGCGGCTCTCTGTCACTGTGTGATGGCTTATGAC 2467  
 Qy 773 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeu 792  
 Db 2468 ACGAGGTACACGGAAAGCAATACATGATGATCCCGAAATATACAGAGAGCTATGAGCA 2527  
 Qy 793 GlySerValAlaMetGlnAlaGluValPheProSerGluProAsnArgLeuLeuLeu 812  
 Db 2528 GGATCTGTAGCCCTGATGTGAGAAAGTCCCAATGAGCTTACACCGCTGCTTATTCCTC 2587  
 Qy 818 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832  
 Db 2588 CACGGCTTCTGAGAGAGAACGTTCACTTCCACACAAATTTCTGTGTGCTCCAGCTG 2647  
 Qy 833 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 852  
 Db 2648 ATCCGACGAGAAAGCCATACCACTTACATCTTACCCAAAGAGAGATACATCCCTCC 2707  
 Qy 853 ValProGluSerGlyGluHisIleTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 872  
 Db 2708 TCCCGGAGATCCGAGAGAGCATTACGAGGTGAGACGCTGCTGCTTCTGACAGAAACACTG 2767  
 RESULT 5  
 AK050023  
 LOCUS  
 DEFINITION  
 AK050023 3376 bp mRNA linear HTC 05-DEC-2002  
 Mus musculus adult male liver tumor cDNA, RIKEN full-length  
 enriched library, clone: C73003H19 product: DIPETIDYL PEPTIDASE 9  
 homolog (Homo sapiens), full insert sequence.  
 ACCESSION  
 AK050023  
 VERSION  
 AK050023.1 GI:26093756  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE  
 AUTHORS  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 20493374  
 81042159  
 AUTHORS  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
 Suni, N., Ishi, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Kashiwagi, K.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 20530913  
 11076861  
 4  
 AUTHORS  
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishi, Y.,  
 Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,



OY	313	eaigtYrProhlysrThrglYtrrAlaAspProlysaIaThrPheIysMetSerGluIle	333
Db	1141	CCGGTACCCACGGACAGCAGCAGACAGAACCCCAAGATTCCCTGAAAGCTGGCTGACCTCA	1200
OY	333	tlleaSpAlaGluGlyArgIleIleAspValIleAspIysGluLeuIleGlnProPheG1	353
Db	1201	GACGGACCATCAGGGCCAAATTCGTGTCAAGCTGCAGAAAGAACTGGTACAGCCATTGAG	1260
OY	353	uileuPheGluGlyValIleGlyTrpIleAlaArgAlaGlyTrpThrProGluGlyIysTy	373
Db	1261	CTCCCTTTTCCCAAGAGGAGTACATGCCCGGGCTGGCTGCAGACCGGAGCCCAATA	1320
OY	373	AlaItrPserIleLeuIleuAspArgSerGlnThrArgLeuGlnIleValIleuIleSerPr	393
Db	1321	TGCCTGGGCCCATGTTCTCGACGCGTCCCAAGACGGCTTCAGCTGTCTCTCTGCCCC	1380
OY	393	oGluLeuPheIleProValIleuAspAspValMetGluArgGluArgLeuIleGluSerVa	413
Db	1381	TGCTCTCTTCAATCCCGGCGGTGTGAGAGAGAGGCGCGGAGCGACGCTCCAGAGCCGT	1440
OY	413	lProAspSerAlaThrProLeuIleIleTyArgGluGluThrAspIleTrpIleAsnI1	433
Db	1441	CCCCAAGAAATGTCAGGCCCTTTGTCTATCTTAAGAAATCACCAATGCTGGATCAACGT	1500
OY	433	eHisAspIlePheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePh	452
Db	1501	CCAGACATCTTCCACCGGTTTCCCAAGGCTGAGGGCCAGACAGACTTTGTCTCTCG	1560
OY	452	eAlaSerGluCysIysThrGlyPheArgHisLeuTyTrpIleHisrIleLeuIysG1	472
Db	1561	TGCCAACAAATGCAAGACTGGCTTCTTGCACCTGTACAGGGTACACAGTGAATTAAAC	1620
OY	472	uSerIysTrpIysArgSerSerGlyIleuProAlaProSerAspPheIysCysProI1	492
Db	1621	CAAGGACTATGACTGACGAGAACCCCTCAGGCTTACAAAGATAGATTAAAGTCCCAT	1680
OY	492	eIysGluGluIleAlaIleThrSerGlyGluTrpGluValLeuGlyArgHisGlySerAs	512
Db	1681	CAAGAGAGAGGTCCGCTCGACACAGTGGGAGTGGGAGGTCTTGTTCGAGCACTGGCTCAA	1740
OY	512	nIleGlnAlaAspGluValArgArgLeuValTyPheGluGluIleHisAspSerProle	532
Db	1741	GATCTGGGTCAACGACGACGACGACMACTGGTGTACTTCAAGGTAAAGAAACACACGCT	1800
OY	532	uGluHisIleLeuTyValIleSerTyValAsnProGlyGluValThrArgLeuThrAs	552
Db	1801	GGAACTACCTCTATGTGTGTGTCAGTCAAGACAGACGAGCGAGATCGTGGCGCTCACCA	1860
OY	552	pArgGlyTyPserHisSerCysIleSerGlnHisCysAspPhePheIleSerIysTy	572
Db	1861	GCTCGGCTTCTCCACACACTGCTCCATAGACGACGCTTGACATGTTGAGAGCTACTA	1920
OY	572	rSerAsnGlnIysAsnProHisCysValSerLeuTyTrpIysLeuSerSerProGluAspAs	592
Db	1921	CAGCAGTGTGAGACACGCCACCTGTGTGATCATGTGTACAAGCTGAGCGGCCCTCATGAGA	1980
OY	592	pProThrCysIysThrIysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPr	612
Db	1981	CCCACTGCACAAAGCAACACCGCTTCTGGGGCAGACATGATGAGGACGCCCAATTGGCCCCC	2040
OY	612	oAspTyThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyArgI	632
Db	2041	AGACTATGTGCCCCCTGAGATCTTCCACTTCACACCCGTCGACACGTCGACGCTTACGG	2100
OY	632	lMetLeuTyTrpIysProHisAspLeuGlnProGlyLeuIleTyTrpProThrValLeuPheI1	652
Db	2101	CATGATCTAACAGCACACACCTCTCAACTGTGGAGGAAGACACCCCATGTGCTCTTGT	2160
OY	652	eTyGlyIysIleProGlnValGlnLeuValAsnAsnArgPheIysGlyValIysTyTrpPheAr	672
Db	2161	CTATGGGGGCCCAACAGGTGCTGTGTGTGAACAATCTTTAAAGGGCATCAATATCTGGCG	2220

Oy		672	gleanstThleuAlaserleuglyTYrValIvalIallexpsnArqGlySerCy	692
Db		2221	GCTAAATACACTGGCATCCCTTGGGCTATGCTGTGTGTATGCATGATGTCGGGCTCCTG	2280
Oy		692	sHsarGlyLeuldsBheGluGlyAlaphelysTyrylsWecGlyGlnileguileas	712
Db		2284	TCAAGGGGGCGCTGACCTTCAGAGGGGCCCTGAATAAATCAATGGCGAGGTGAGATTGA	2340
Oy		712	pAeGInValIGlUGlyleuGIntTyrlEualAserArGYrAspPheIlleAspleAsPAR	732
Db		2341	GGACCAGATGGAAAGCTTGAGTAGACGtGGCTGAGAAATATAGCTTCAATTAAGCTTGAACCG	2400
Oy		732	gvaIGlyIleHiseglyTrpserTyrglygylyTYrLeuSerleuMetAlaleuMetGlnar	752
Db		2401	AGTGGCCATCCATGCTGCTGCTCTAAGCGGCGCTTCTCTCACTCATGGGCGCTCATCCAA	2460
Oy		752	gsErAsPIlePheArGyValAlaleaGlYAlaPrOvalThrleuTrpIIepHeTyras	772
Db		2461	CCACAAGTGTTCACAGTACCAATGGCGGCGCTCCGTCACTGTGTGATGGCGTTATGA	2520
Oy		772	pTHGIlyTyThrGlyuarGtyrMetGlyHisProAspGlnAsndJungInglyTYrTyre	792
Db		2521	CACAGGATCACGGAACGATACATGATGTCCCGAAAATAACGACGAAAGCTATGAGCC	2580
Oy		792	uGlySerValAlaMerGlnlaIgulyrsPheProSerGluProAsnArgleuleuLeule	812
Db		2581	AGGGTCTGTAGCCCTGCATGTGGAGAAGCGCCCAATGAGACCTPAACCGCTGATTATCT	2640
Oy		812	uHisGlyPhleleuAspGluAsnyValHisPheAlaHisThrSerlleleuSeurPhele	832
Db		2641	CCAGGCGCTCTCGACAGAAACGTTCACTTCTTCCAACAATTTCTGTGGTCTCCAGCT	2700
Oy		832	vUaIarAlaGlyLyseProTyraSpleuGlnIleTyProGlngluArghIsSerIleAr	852
Db		2701	GATCCGAGCAGAAAGCCATACACAGCTTCATGATCTTCAACGAGAGACATACATCCG	2760
Oy		852	gVAlprGluSerGlygluHisTyrgIuleuHisleuLeuHisTyrlleuGngluAnle	872
Db		2761	CTGCCGGAGTCCGAGACATTACGAGGTGACCGCTCTCACTTCTGACAGAAACCT	2820
Oy		872 u	872	
Db		2821 G	2821	
RESULT 6		0	'	
BMS57438				
LOCUS			1042 bp	mRNA linear EST 20-FEB-2002
DEFINITION			AGNCOURT 6578992 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466759	
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM			Homo sapiens (human)	
TITLE			Homo sapiens	
AUTHORS			Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
JOURNAL			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
COMMENT			NIH-MGC htcp://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strauberg, Ph.D. Email: rsgabbs@emall.nih.gov Tissue Procurement: DCTD/DRP CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.jnl.gov plate: LCM1967 row: k column: 16 High quality sequence stop: 697.	
FEATURES				
SOURCE			I..1042	

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:546759"
/risue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 41"
/notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT      279 a      223 c      214 g      303 t      3 others
ORIGIN
Alignment Scores:
Pred. No.:      5,23e-178      Length:      1042
Score:          1693.50      Matches:      324
Percent Similarity: 93.97%      Conservative: 3
Best Local Similarity: 93.10%      Mismatches: 15
Query Match:     36.03%      Indels:      6
DB:              12      Gaps:      1
US-10-070-464-1 (1-882) x BM557438 (1-1042)
OY 504 GluValLeuGlyAyrHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyr 523
Db 1 GAAGTTCTTGGCGCGCATGATCATTAATCCAAAGTTGATGAATGAGAAAGCGGTGATAT 60
OY 524 PheGluGlyThrThyAspSerProLeuGluHisIleLeuTyrValValSerTyrValAsn 543
Db 61 TTGGAAGCACCAAGACTCCCTTTAGACATCACTGTAGTACGATCACTTAAT 120
OY 544 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 563
Db 121 CCTGAGAGGCTACAAAGCTGACGACCGTGGCTACTCATCTTGGCGATCACTG 180
OY 564 HisCysAspPhePheIleSerIleTyrSerAsnGlnIleAsnProHisCysValSerLeu 583
Db 181 CACTGTGACTCTTTATAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 240
OY 584 TyrIleAspSerSerProGluAspAspProThrCysIleThrThyGluPheThrAlaThr 603
Db 241 TACAAGCTATCAAGCTCGAAGATGACCAACTTCGAAAAAGAAATTGGGCGCAC 300
OY 604 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 623
Db 301 ATTTGGATTACAGCAGGTCCTCTCTGACTATCTCCCTCCGAAATTTCTCTTTGAA 360
OY 624 SerThrThrGlyPheThrLeuTyrGlyIleCysLeuTyrIlePheHisAspLeuGlnProGly 643
Db 361 AGTACTACTGATTTACATTGATGAGATGCTCTCAAGCCCTCATGATCTACAGCTGGA 420
OY 644 LysIleTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 663
Db 421 AAGAAATATCCACTGCTGCTGCTCATATATGAGTGGCTCCAGTGGTGTGATATAT 480
OY 664 ArgPheIleGlyValIleTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 683
Db 481 CGGTTAAAGGAGTCAGATATTTCCGCTTGAATACCTTACGCTCTCTAGGTTATGTGTT 540
OY 684 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLeuPheGluGlyAlaPheIle 703
Db 541 GTAGATAGACAAACAGGGGATCCCTGACCAAGGGCTTAAATTGAAAGGCCCTTTTANA 600
OY 704 TyrIleMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 723
Db 601 TATTAATAGGTCAAATAGAAATTAACATCAGGTCGAGAGGATCTCAATATCTTACTCT 660
OY 724 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrGlyTyr 743

```

```

Db 661 CGATATGATTTTATTGACTTATGATCTGTGGGCATCCACGGCTGGTCCATGAGAGATAC 720
OY 744 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 763
Db 721 CTCTCCCGATGCGCATTAATGACAGAGTCAGATATCTTCAGGGTTGCTATTCCTGGGGCC 780
OY 764 ProValThrLeuTyrIlePheThrAspThrGlyTyrThrGluValArgTyrMetGlyHisPro 783
Db 781 CCAGTCACTCTGTGATCTCTTATATACAGATACAGGATACAGGTCGTTATATAGGATCCCT 840
OY 784 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluIlePhePro 803
Db 841 GACCGAATGAAACAGGGCATATTACTTATGATCTGTGGCATGCAAGCAAGAAAGTCCCC 900
OY 804 SerGluProAsnArgLeuLeuLeuLeuHisGly-PheLeuAspGluAsnVal-HisPheA 823
Db 901 TCTGAACCAATCGATTACTGCTCTTACATGCTTTTCTCGATGAGAAATGTCATTTTG 960
OY 823 IaHisThrSer-IleLeuLeuSerPheLeuValAlaGlyIleProTyrAspLeu--- 841
Db 981 CACATACCCAGAGATATTACTGAGTTTNTTGTGAGAGGGCTGAAAAAGCCATATGAT 1020
OY 842 -----GlnIleTyrPro 845
Db 1021 TTACACCAATCTATCT 1038
RESULT 7
BO675006      957 bp      mRNA      linear      EST 15-JUL-2002
LOCUS      AGENCOURT_8303592 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6274622
DEFINITION      5', mRNA sequence.
ACCESSION      BO675006
VERSION      BO675006.1 GI:21785840
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 957)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bde-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM2456 row: h column: 15
High quality sequence start: 4
High quality sequence stop: 624.
FEATURES
location/Qualifiers
1..957
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6274622"
/risue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 102"
/notes="Organ: salivary gland; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Library constructed
by Ling Hong in the Laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      257 a      206 c      228 g      266 t

```



## ORIGIN

## Alignment Scores:

Pred. No.: 3,95e-161 Length: 957  
 Score: 1542.00 Matches: 298  
 Percent Similarity: 95.24% Conservative: 2  
 Best Local Similarity: 94.60% Mismatches: 10  
 Query Match: 32.81% Indels: 5  
 DB: 13 Gaps: 0

US-10-070-464-1 (1-882) x BQ675006 (1-957)

QY 474 LysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLys 493  
 DB 12 AAATATAACGATCCAGTGGTGGCTGCTCCAGGATTTCAAGTGTCTCTATCAAA 71  
 QY 494 GlnGluAlaAlaIleThrSerGlyGluTrrGluValLeuGlyArgHisGlySerAsnIle 513  
 DB 72 GAGGAGATGACAAATTACCAAGTGGTGAATGGGAAGTCTTGCGCGCATGGATCTAAATATC 131  
 QY 514 GlnValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlu 533  
 DB 132 CAAGTTGATGAAGTCAAGAGGCTGGTATATTTGAAGGACCAAGAGCTCCCTTTAGAG 191  
 QY 534 HisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArg 553  
 DB 192 CATCACCTGTACCTAGTACGATTACGTAAATCTCGAAGAGGTGACAAAGGCTGACCTGACCGT 251  
 QY 554 GlyTyrSerHisSerCysLysSerGlnHisCysAspPheIleSerLysTyrSer 573  
 DB 252 GGCTACTACATCTTGCTGCTGCATCAGTCAGCTGAGCTCTTTATTAAGTAAATAGT 311  
 QY 574 AsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerProGluAspAspPro 593  
 DB 312 AACCAAGAGATCCACACTGTGTGCTCTTACAGATCAAGCTCCTGAAGATGACCCA 371  
 QY 594 ThrCysLysThrLysGluPheTrrPAlaThrIleLeuAspSerLysGlyProLeuAspAsp 613  
 DB 372 ACTTGGAAAAAAGGAATTTTGGCCACCATTTTGGATTGACAGGCTCTCTTCTGAC 431  
 QY 614 TyrThrProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 633  
 DB 432 TAACTCTCCAGAAATTTTCTTTTGAAGTACATGATTTACATTTATGGGATG 491  
 QY 634 LeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyr 653  
 DB 492 CTCCTACAGCCTCATGATCTACAGCCTGGAAGAAATATCTACTGCGCTTCATATAT 551  
 QY 654 GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 673  
 DB 552 GGTGGTCTCCAGGTGACGTTGGTGAATATCGGTTTAAAGAGATCAAGTATTTCCGCTTG 611  
 QY 674 AsnThr-LeuAlaSerLeuGlyTyrValValValAlaLeuAsp-AsnArgLysSerCysH 693  
 DB 612 AATACCCCTAGCCTCTCTAGTATGTGTGTAGTATGACAAACAGGGGATCTCTGTC 671  
 QY 693 IsetrGlyLeuLysPheGlyGlyAla-PheLysTyrLysMet-GlyGlnIleGluIleAs 712  
 DB 672 ACCGAGGGCTTAAATTTGAAGGCGCCTTTAATATATTAATGAGGCAATATAGAAATTTGA 731  
 QY 712 PArgGlnValGlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleLeuAspAsp 732  
 DB 732 CGATTCAGGGGAAAGACTCCAAATATCTAGCTTCTCGAATATGATTTCAATGATTAAGTGC 791  
 QY 732 GValGlyIleHisGlyTrrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752  
 DB 792 TGTGGGCATCCACGGCTGGTCTTATGAGAGATCCCTCTCCCTGAAGCATTTATGACAG 851  
 QY 752 GSerAspIlePheArgValAlaIleAlaGlyAlaProValThr-LeuTrrIlePheTyr 772  
 DB 852 GTCCAGATATCTTCCAGGTTGCTATGTCTGGGGGCCAGTCACATCCGTTGATCTTCATG 911  
 QY 772 spThrGlyTyrThrGluArgTyrMetGlyHisPro 783

DB 912 ATACAGATACCCGGAACGTTATATGAGGCCCCC 946

RESULT 8  
 LOCUS BQ675260 910 bp mRNA linear EST 15-JUL-2002  
 DEFINITION AGENCOUNT 8354972 NIH\_MGC\_102 Homo sapiens cDNA clone IMAGE:6275478  
 5', mRNA sequence.  
 ACCESSION BQ675260  
 VERSION BQ675260.1 GI:21786094  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 910)  
 NIH-MGC http://mgs.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 CONTACT: Robert Strausberg, Ph.D.  
 COMMENT Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L10C2458 row: 1 column: 07  
 High quality sequence stop: 618.  
 Location/Qualifiers

## FEATURES

## source

1..910  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6275478"  
 /issue\_type="epidermoid carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1b="NIH\_MGC\_102"  
 /note="Organ: salivary gland; Vector: pOT87; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Library constructed  
 by Ling Hong in the laboratory of Gerald M. Rubin  
 (University of California, Berkeley) using ZAP-CDNA  
 synthesis kit (Stratagene) and Superscript II RT (Life  
 Technologies). Note: this is a NIH\_MGC library."

## BASE COUNT

248 a 186 c 212 g 262 t 2 others

## ORIGIN

Alignment Scores:  
 Pred. No.: 6.9e-161 Length: 910  
 Score: 1539.50 Matches: 293  
 Percent Similarity: 97.04% Conservative: 2  
 Best Local Similarity: 96.38% Mismatches: 5  
 Query Match: 32.76% Indels: 4  
 DB: 13 Gaps: 1

US-10-070-464-1 (1-882) x BQ675260 (1-910)

QY 475 TyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGlu 494  
 DB 2 TATTAACGATTCAGTGGTGGCTGCTCCAAAGTATTCAAAGTCTTATCAAAAG 61  
 QY 495 GlnIleAlaIleThrSerGlyGluTrrGluValLeuGlyArgHisGlySerAsnIleGln 514  
 DB 62 GAGATGCAATTAACAGTGTGAATGGGAAGTCTTGCGCGCATGGATCTAATATCCA 121  
 QY 515 ValAspGluValArgArgLeuValTyrPheGlyGlyAlaLeuAspSerProLeuGluHis 534  
 DB 122 GTTGATGAAGTCAAGAGCTGGTATATTTGAAGGACCAAGAGATCCCTTAAAGCAT 181  
 QY 535 HisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGly 554



```

Db      |||||
182  CACCTGTACGTAGTACGTACGTAAATCCTGGAGAGGTGACAAAGCTGACCTGACCGTGGC 241
Qy      |||||
555  TYSerHisSerCyCyAlleserGlnHisCyAspPhePhe11eser1ysrSerAan 574
Db      |||||
242  TACTACATCTTCTGTCAGTACGTACGACTGTGACTCTTTATTAAGTAAGTACTATAC 301
Qy      |||||
575  GlnLysAsnProHisCyValSerLeuTyTyLysLeuSerSerProGluAspAspProtr 594
Db      |||||
302  CAGAAATCCACACCTGTGTCTTCAACCTTCAAGTCTTAAGTACCACT 361
Qy      |||||
595  CysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTy 614
Db      |||||
362  TCCAAAACAAGAAATTTGGGCCACATTTGATTGACGAGTCTCTTCTGACTAT 421
Qy      |||||
615  ThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyTyGlyMetLeu 634
Db      |||||
422  ACTCTCCAGAAATTTCTTTTGAATACACTACGATTAACATGATGAGGAGCTC 481
Qy      |||||
635  TyTyLysProHisAspLeuGlnProGlyLysTyTyProThrValLeuPheIleTyTyGly 654
Db      |||||
482  TCAACACCTCATGATTTACAGCTGGAAGAAATATCTTACTGTCTTCATATATGAT 541
Qy      |||||
655  GlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyTyPheArgLeuAsn 674
Db      |||||
542  GGTCTTCAGGTGACGTGTGATGAAATATCGATTTAAAGAGTCAAGTATTTCCGCTTGAAT 601
Qy      |||||
675  ThrLeuAlaSerLeuGlyTyTyValValValIleAspAsnArgGlySerCyHisAlaG 694
Db      |||||
602  ACCCTAGCTCTCTAGGTATGATGTTGATGATGACAAACAGGGGATCTGTCTACCA 661
Qy      |||||
695  GlyLeuLysPheGluGlnValaPheLysTyTyLysMetGlyGlnIleGluIleAspAspGln 714
Db      |||||
662  GGGCTTAATTTGAAAGCCCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 721
Qy      |||||
715  ValGluGlyLeuGlnIleLeuAlaSerArgTyTyAspPheIleAspLeuAspArgValGly 734
Db      |||||
722  GTGAAGAAGACTCCATATATCTAGCTTCTCGATATGATTTTCACTTACATGATCGTGGGC 781
Qy      |||||
735  IleHisGlyTrpSerTyTyGlyTyTyLysSerLeuMetAla-LeuMetGlnArgSerA 754
Db      |||||
762  ATCCACGGCTGATCCATAGAGATACCTCTCCGATGATGACATTAAACAGAGCA 841
Qy      |||||
754  P-IlePheArgValAlaIleAlaGlyAlaProValThrLeuTrp---Ile-PheTyTyAsp 772
Db      |||||
842  TTAATCTTCAAGGTGTCTATTTGCTGGGCCCCCAAGTCAATTCGGGGATCTTTCTATGAT 901
Qy      |||||
773  ThrGly 774
Db      |||||
902  ACAGGG 907

```

```

RESULT 9
LOCUS   BI084090 1041 bp mRNA linear EST 20-JUN-2001
DEFINITION 602869453p1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5013996 5',
ACCESSION BI084090
VERSION   BI084090.1 GI:14502420
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1041)
NIH-MGC http://mgi.mgi.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-riemail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLC1820 row: 5 column: 13  
 High quality sequence stop: 847.  
 Location/Qualifiers

1.1041

```

/organism="Homo sapiens"
/mol_type="RNA"
/db_xref="taxon:9606"
/clone="IMAGE:5013996"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 102"
/note="Torgan: salivary gland, Vector: pOT87, Site_1: XhoI;
Site_2: EcoRI, cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH-MGC Library."

```

BASE COUNT 280 a 227 c 242 g 292 t

#### Alignment Scores:

```

Pred. No.: 5.5e-157 Length: 1041
Score: 1505.50 Matches: 311
Percent Similarity: 89.04% Conservative: 6
Best Local Similarity: 87.36% Mismatches: 20
Query Match: 32.03% Indels: 21
DB: 12 Gaps: 3

```

US-10-070-464-1 (1-882) x BI084090 (1-1041)

```

Qy      |||||
474  LysTyTyAspSerSerGlyGlyLeuProAlaProSerAspPheLysCyProIleLys 493
Db      |||||
3  AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 62
Qy      |||||
494  GluGluIleAlaIleThrSerGlyGlyTrpGluValLeuGlyArgHisGlySerAanIle 513
Db      |||||
63  GAGGAGATGACAAATTAACAGTGTAAATGGAAAGTCTTGGCGGATGATTAATATC 122
Qy      |||||
514  GlnValaAspGluValArgArgLeuValTyTyPheGluGlyThrLysAspSerProLeuGlu 533
Db      |||||
123  CAAGTGTAGAGTCAAGAGCTGTATATTTTGAAGGACCAAGACCTCCCTTTAGAG 182
Qy      |||||
534  HisHisLeuTyTyValValSerTyTyValaAspProGlyGluValThrArgLeuThrAspArg 553
Db      |||||
183  CATCACTGTACGTAGTACGTAAATCTCGAGAGAGTACAAAGCTGACCTGACCT 242
Qy      |||||
554  GlyTyTySerHisSerCyCyAlleserGlnHisCyAspPhePhe11eser1ysrSer 573
Db      |||||
243  GGTACTACATCTTCTGTCAGTACGTACGACTGTGACTCTTTATTAAGTAAGTACT 302
Qy      |||||
574  AsnGlnLysAsnProHisCyValSerLeuTyTyLysLeuSerSerProGluAspAspPro 593
Db      |||||
303  AACCAAGAAATCCACACCTGTGTCTTCAACCTTCAAGTCTTAAGTACCACT 362
Qy      |||||
594  ThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp 613
Db      |||||
363  ACTTCAAAAACAAGAAATTTGGGCCACATTTGATTGACGAGTCTCTTCTGACT 422
Qy      |||||
614  TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyTyGlyMet 633
Db      |||||
423  TATACTCTCCAGAAATTTCTTTTGAATACACTACGATTAACATGATGAGGAGAT 482
Qy      |||||
634  LeuTyTyLysProHisAspLeuGlnProGlyLysTyTyProThrValLeuPheIleTyTy 653
Db      |||||
483  CTCTACAAAGCTCATGATTTACAGCTGGAAGAAATATCTTACTGTCTTCATATAT 542
Qy      |||||
654  GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGly--ValLysTyTyPheArgL 673

```

```

Db      543 GGTGCTCCAGGTCGACGTTGGTGAATATCGGTTTAACAGGACGATCAAGTATTTCGGCT 602
Qy      673 euAnsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCysH 693
Db      603 TGAATACCCCTAGCTCTCTAGTATGTGGTTGTAGTATAGACCAAGGGGATCTCTGTC 662
Qy      693 iAArgGlyLeuLysPheGlyValAlaPheLysTyrLysMetGlyGlnIleGlnIleAspA 713
Db      663 ACCGAGGGCTTAAATTTGAAGGCGCTTTAA-TATCAATGGCTCATATA-GAAATTGACG 720
Qy      713 spGlnValGluGlyLeuGln-TyrLeuAlaSerArgTyrAspPheIle-AspLeuAspArg 732
Db      721 ATACGCTGAAAGACCTCCAAATTAATCTAGCTTCGATATGATTAATTCATTGACTTAAGTC 780
Qy      732 gValGlyIleHisGlyTyrPheTyrGly-GlyTyrLeuSerLeu-MetAlaLeuMetGln 751
Db      781 TGTGGGCAATCCACGGGTGGTCTATGAGAGGATACCTCTCCGTAATGGCATTAATGCGAG 840
Qy      752 ArgSerAsp-IlePheArgValAlaIleAla-GlyAlaProValIThrLeuTPIlePhe- 770
Db      841 AGGTCAAGATACCTTCACAGGTCGCAATAGCTGGGGGCGCCAGTCACTCTGTGATCTTCT 900
Qy      771 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyr 790
Db      901 TATGATACAGATACACGAGACCTTATATGGGTCACTGACCTGAACAGA---TGAACGGGTCA 957
Qy      791 TyrLeuGlySerValAlaMetGlnAlaGluLysPhePro----- 803
Db      958 ACTTCAGATCTGTG-----CCTGCAGCGAAATTCCTTGCATATCGTAAATGCTCA 1011
Qy      804 SerGluProAsnArgLeuLeuLeuHis 813
Db      1012 ATGCTTCCGGCGAAATTCCTGATACAC 1041

RESULT 10      968 bp      mRNA      linear      EST 15-JUL-2002
BO671635
LOCUS      BO671635
DEFINITION  5', mRNA sequence.
ACCESSION  BO671635.1 GI:21782469
VERSION    BO671635.1
KEYWORDS
SOURCE
ORGANISM   Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabds-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LDCM2407 row: a column: 22
High quality sequence stop: 630.
Location/Qualifiers
1..968
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6255645"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="PHIOB (phage-resistant)"
/clone_1ib="NIH_MGC_102"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the

```

following 5' adaptor: GGCACAG (G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

## ALIGNMENT SCORES:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1..35e-148	968	1430.00	280	6	29	11	2
Percent Similarity:		87.73%					
Best Local Similarity:		85.89%					
Query Match:		30.43%					
DB:		13					

US-10-070-464-1 (1-882) x BO671635 (1-968)

```

Qy      475 TyrLysArgSerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGlu 494
Db      3 TATAAACGATCCAGTGTGGCTGCTGCTCCAAAGATTTCAAGTCTTCAAAAGAG 62
Qy      495 GlnIleAlaIleThrSerGlyGluTyrGluValIleGlyArgHisGlySerAsnIleGln 514
Db      63 GAGATAGCAATTAACAGTGTGAATGGAAATTTTGGCCGCGATGATCTAATATCCAA 122
Qy      515 ValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHis 534
Db      123 GTTGAATGAAGTACGAAGCGTGGTATATTTTGAAGGACCAAAAGACTCCCTTTAGAGCAT 182
Qy      535 HisLeuTyrValValSerTyrValAsnProGlyGluValIThrArgLeuThrAspArgGly 554
Db      183 CACCTGATAGTACGATGATTAATCTCGAGAGGTACCAAGGCTGACTACCTGGCC 242
Qy      555 TyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsn 574
Db      243 TACTACATCTTCTTGCTGATCAGTACGACCTGACCTTTTATATGTAAGTATAGTAAAC 302
Qy      575 GlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThr 594
Db      303 CAGAAAGATCCACACGTGTGTGTCCTTTTACAAGCTATCAAGTCTGTAAGATACCAACT 362
Qy      595 CysLysThrLysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyr 614
Db      363 TCCAAACCAAGAAATTTTGGCCACACATTTTGGATTTCAGAGGTCTCTCTCGTACTAT 422
Qy      615 ThrProProGlnIlePheSerPheGlySerThrThrGlyPheThrLeuTyrGlyMetLeu 634
Db      423 ACTCTCCAGAAATTTCTCTTTTAAAGTACTGATTTTACATTTGATGGAGTGC 482
Qy      635 TyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGly 654
Db      483 TCAAGAGCTCATGATCTACAGCTCGAAGAAATATCTCACTGTCTGTCTATATATGAT 542
Qy      655 GlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsn 674
Db      543 GGTCTCTAGGTGCAAGTGTGTAATAGGTTTAAAGAGTCAAGATATTCGGCTTGAT 602
Qy      675 ThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArg 694
Db      603 ACCCTAGCTCTCTAGTATATGTGTTAGTATAGACCAACAGGAGATCTGTACACGA 662
Qy      696 GlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGln 714
Db      663 GGGCTTAATTTGAAGCGCTTTAATATATAAATGGGTCAAAATGAATATGACATCAG 722
Qy      715 ValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg-ValG 734
Db      723 GGGGAAGACCTCAATATCTAGCTTCTGATATGATTTTCATTGACTTGAATCGAGTGG 782
Qy      734 TTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 753
Db      783 CATCACCGCTGGGCGCTATGGAGGATACCTCTCTCCCTGATGGATTTAATGGCGGAG 842

```

Qy 754 AspiepheargValAlaIlelela-----GIYAlaProvalThlrleuphlephe 770  
Db 843 TCCCAATATCTTCCCGGGGCTTATAGCCTCGGGGAGCCCACTCCCGCC----- 893  
Qy 771 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyr 790  
Db 894 -----GGGGGATACCTCCCAAGCAACCGGAGACCCGGGAGCGTTCGGCGGTGGTGC 947  
Qy 791 TyrleuGlySer 794  
Db 948 CCCCTCGGTGCC 959

RESULT 11  
BX372276/c 910 bp mRNA linear EST 08-MAY-2003  
LOCUS BX372276 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL005YD02 3-PRIME, mRNA sequence.  
ACCESSION BX372276 GI:30448117  
VERSION BX372276.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 910)  
Li W.B., Gruber C., Jessee J. and Polayes D.  
Full-length cDNA libraries and normalization  
Unpublished  
JOURNAL  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to this cluster, see  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BA1043ZH06\_CS04080\_1&cluster=7542.r.  
Contact: Feng Liang Email: fliang@lifetech.com URL: [http://fulllength.invitrogen.com/InvitrogenCorporation1600ParadiseAvenueGenoscopeSequenceID%3DCS0BA1043ZH06\\_CS04080\\_1.](http://fulllength.invitrogen.com/InvitrogenCorporation1600ParadiseAvenueGenoscopeSequenceID%3DCS0BA1043ZH06_CS04080_1.Location/Qualifiers)  
FEATURES  
source  
1..910  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL005YD02"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 251 a 193 c 176 g 288 t 2 others  
ORIGIN

Alignment Scores:  
Pred. No.: 2,62e-148 Length: 910  
Score: 1427.00 Matches: 293  
Percent Similarity: 96.38% Conservative: 0  
Best Local Similarity: 96.38% Mismatches: 11  
Query Match: 30.36% Indels: 5  
Gaps: 13

US-10-070-464-1 (1-882) x BX372276 (1-910)

Qy 305 LeuGluThrArgGlyAlaAspSerPheArgTyrProlySThrGlyThraAlaAspProlys 324  
Db 907 TTGGAAACAAAGAGGAGGATTCATTCGCG-TATCTAAACAGGACGCAATCTTAA 849  
Qy 325 ValThrPheIysMetSerGluIleuMetIleAspAlaGluGlyAlaIleleleAspValIle 344

Db 848 GTCACTTTTAAGATGTCAAGAAATATGATGTATGCTGAAAGAAAGATCATATGTCANA 789  
Qy 345 AspylGluLeuIleGlnProPheGluIleleuPheGluGlyValGluTyrIleAlaArg 364  
Db 788 GATAAGAACTAATTCACCTTTTGAGATTCATAT-GAAGNAGTTGATATATTCACAGA 731  
Qy 365 AlaGlyTPThrProGluGlyIleTyrAlaIlePheSerIleleuLeuAspArgSerGlnThr 384  
Db 730 GCTGATGACTCTCGAGAGAAATATGCTGTGTCATCTCATCTGATGATGCTCCAGACT 671  
Qy 385 ArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGluAspAspValMet 404  
Db 670 CGCCTACAGATAGTGTATCTCATCTCACTAATATTATTCNCAGTGAAGATGATGTATAG 611  
Qy 405 GluArgGlnArgLeuIleGluSerValProAspSerValThrProLeuIleIleTyrGlu 424  
Db 610 GAAAGGCGAGACATCATGATGACAGTCCCTATCTCGAGCCCACTAATATCATGAA 551  
Qy 425 GluThrThrAspIleTPThrIleAsnIleHisAspIlePheHisValPheProGlnSerHis 444  
Db 550 GAAACACAGACATCTGATTAATATCCATGACATCTTCATGTTTCCCAAGTCCAC 491  
Qy 445 GluGluGluIleGluPheIlePheIleAspGluCysIleTyrGlyPheArgHisIleTyr 464  
Db 490 GAAGAGAAATTTGATTTATTTTTCCTGCTGAATGCAAAACAGTTTCGTCATTTATAC 431  
Qy 465 LysIleThrSerIleleuLeuGluSerIleTyrIleAspArgSerGlyGlyLeuProAla 484  
Db 430 AAAATTACATCATTTTAAAGAAAGCAAAATATAAACGATCCAGTGCGGCTGCTGCT 371  
Qy 485 ProSerAspPheIysCysProIleleuGluGluIleAlaIleThrSerGlyGluTyrGlu 504  
Db 370 CCAAGTGAATTCAGATGCTCTCATCAAAAGAGAGATGAGCAATTCACGATGTAATGGAA 311  
Qy 505 ValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrPhe 524  
Db 310 GTTCTGCGCCGCGATGATCTAATATCAAGTTGATGAGACAGAGCGTGATATTTT 251  
Qy 525 GluGlyThrIleAspSerProLeuGluHisIleIleTyrValIleSerTyrValAsnPro 544  
Db 250 GAAGGACACAAAGACTCCCTTTAGAGATCATCGTACGATGATGATGATGATGATGAT 191  
Qy 545 GlyIleValIleThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHis 564  
Db 190 GGAGAGTGAGACAGGCTGACGACCGTGGCTACACATTTCTGTCATCATGTCAGCAC 131  
Qy 565 CysAspPhePheIleSerIleTyrSerAsnGlnIleAsnProHisCysValSerIleTyr 584  
Db 130 TGTGACTCTTTTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 71  
Qy 585 LysLeuSerSerProGluAspAspProThrCysIleTyrIleGluPheTyrAlaThrIle 604  
Db 70 AAGCATATCAAGTCTGAAAGATGACCAAC-TGCAAAAACAAAGAAATTTGGCCACCAT- 13  
Qy 605 LeuAspSerAla 608  
Db 12 TTGATTCAGCA 1

RESULT 12  
BX390898 919 bp mRNA linear EST 13-MAY-2003  
LOCUS BX390898 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL005YD02 5-PRIME, mRNA sequence.  
ACCESSION BX390898  
VERSION BX390898.1 GI:30607432  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 919)  
Li W.B., Gruber C., Jessee J. and Polayes D.





Qy	538	valValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis	557
Db	182	GTGTGAACACTACAAAGATTCCTGGAGAAGTAAGCACTGACAGAAACGTGGTAACTCTCAT	241
Qy	558	SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn	577
Db	242	GCCTCTGTGTCCAGCCAGGATTTGTGACATGTTCAATCAGCAATGACGACATCAGAAAGAC	301
Qy	578	ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr	597
Db	302	CCACACTGTGTCTCCCTTTACCCGGCTGACAGGACATGAAGATGACAGCTCAAAAGACA	361
Qy	598	LysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro	617
Db	362	AAGGAATTCGGGCTACAAATTTGGATTCAGCAGGCCCTTCCTGATTGACTTCCCCA	421
Qy	618	GluIlePheSerPheGluSerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysPro	637
Db	422	GAAAGTGTCTCTTTGAGAGCTCCACGGGGTTTAACTCTTAAGGATGATGACAAACT	481
Qy	638	HisAspLeuGlnProGlyLysLysTyrProThrValIleuPheIleTyrGlyGlyProGln	657
Db	482	CACATCTCGCAACCTGGAAAGAGAACCTTACTGTGATTTTATCTTAATGAGGCCCTCAG	541
Qy	658	ValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAla	677
Db	542	GTGCAGCTAGTGAACAATCGATTTAAAGGACTCAATATTTCCGATTGAACACTTGGCC	601
Qy	678	SerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLys	697
Db	602	TCTTTAGGCTACGTGTGTCTGTATATGACAAACGGGGGTCTGTGCACCGAGGGCTGAAG	661
Qy	698	PheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGlu-Gl	717
Db	662	TTTGAAGGAGCCTTTAAATCAAAATGGGACAAATTAAGAAATTGATGACCAAGTGAAGG	721
Qy	717	LysGlnTyrIleuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGly	737
Db	722	TCTGACACTACTTTGGCGTTCAGATAGACTCTTAATTGGCTTCCGGGTGGGACTTCATGG	781
Qy	737	YTrpSerTyrGlyGlyTyr---LeuSerLeuMetAlaLeuMetGlnAspSer-AspIleP	756
Db	782	GTGGCCCATGAGAGGCGTAACTCTCTCTTAATGGCGCTTTAAATCCGAGTCAAGATTACT	841
Qy	756	heArgValAlaIleAlaGlyAla-ProValThrLeuTrp---IlePheTyrAspThrGly	774
Db	842	TCAGGGTTCATCCCGGATCCCGGTAACCGCTGGGAGATTCTTATAGACACGGGT	901
Qy	775	-TyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlyLys	794
Db	902	ATACACGAGAGCGCTTAATGCGGCCCCCGGGGACACACGACGAGGTATTA--CTGGGTCA	959
Qy	794	rValAlaMetGlnAlaGluLysPhe 802	
Db	960	G--G--GGCATGCAAGCGAGAAAGTTC 981	
RESULT 15			
AL040398			
LOCUS	AL040398	753 bp	mRNA linear EST 29-FEB-2000
DEFINITION	DKFZp344A0714.r1.434 (synonym: hres3) Homo sapiens cDNA clone		
ACCESSION	DKFZp344A0714.5', mRNA sequence.		
VERSION	AL040398		
KEYWORDS	AL040398.1	GI:5409350	
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.		
TITLE	EST (Koehrer, et al.)		
JOURNAL	Unpublished		
COMMENT	Contact: Koehrer K		

**MPIS**  
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
 This is the 5' sequence of the clone insert  
 clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de,  
 sequenced by BMFZ (Biomedical Research Center at the Charité,  
 Berlin/Germany) within the CDNA sequencing consortium of the German  
 Genome Project.  
 No sl sequence available.  
 This clone (DKFZP434A0714) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Neuherberg 6, 14059  
 Berlin-Charlottenburg; GERMANY; Email: clone@rzpd.de.  
 Location/Qualifiers  
 1..753

BASE COUNT	218 a	150 c	170 g	214 t	1 others
ORIGIN	/clone_1ib="434 (synonym: htees3)" /note="Vector: pSPori1; Site_1: NotI; Site_2: SalI"				
Alignment Scores:					
Pred. No.:	4,68e-139	Length:	753		
Score:	1343.00	Matches:	249		
Percent Similarity:	99.60%	Conservative:	0		
Best Local Similarity:	99.60%	Mismatches:	1		
Query Match:	28.57%	Indels:	0		
DB:	9	Gaps:	0		
US-10-070-464-1 (1-982) x AL040398 (1-753)					
QY	466	11ethSer11eLeu5GluSer1yfrYrYsAArgSerSerG1yG1yLeuProA1aPro	489		
Db	1	ATTACATCTATTATTAAGGAAGCAATAATTAAGATCACTGATGGAGCTGCTGCTCA	60		
QY	486	SeAspPheLyScySPro11eLyG1uG1u11eA1a11eThSerG1yG1uTrpG1uA1	505		
Db	64	AGGATTTCAAGTCGTCATCAAGAGAGAGATACCAATTACATGTCGATGGAGATT	120		
QY	506	LeuG1yA1gH1sG1ySerAsn11eG1nA1aSpG1uA1aArgArgLeuA1yTrpHeG1u	525		
Db	121	CTTGGCGGCAAGATCTCAATATCAACATTCATGAAGTCAGAGAGCTGGATATTTTGA	180		
QY	526	G1yTrYrYsAspSerProLeuG1nH1s1eLeuYrYsA1aSerYrYsA1aSpProG1y	545		
Db	181	GGGACCAAGACTCTCCCTTTAGAGCATCACTGTACGTAAGTACGTAAATCCCTGGA	240		
QY	546	G1uA1a1ThrArgLeuThrAspArgG1yTrSerH1sSerCyS11eSerG1nH1sCyS	565		
Db	241	GAGGAGACAAGCTGATCGACCGTGGCTACTCACATTCCTTGCTGCATCATGACGACTGT	300		
QY	566	AspPhePhe11eSerLyTrSerAsnG1nLyAsnProH1sCyS1aSerLyTrYrYs	585		
Db	301	GACTTCCTTATATAGTAAGTATAGTAACACAGAAAGATCCACACTGTGTCTCTTACAG	360		
QY	586	LeuSerSerProG1aAspAspProThCyS1yArThrLyG1uG1uPheTrpA1aTr11eLeu	605		
Db	361	CTATCAAGTCCTGAGATGACCCCACTTGCAAAACAAAGAAATTTTGCCGACCATTTTG	420		
QY	606	AspSerA1aG1yProLeuProAspTrYrThProG1u11ePheSerPheG1uSerThr	625		
Db	421	GATTGAGAGGCTCTCTCTGACTATATCTCTCCCAAAATTTCTCTTTTGAAGACT	480		
QY	626	ThrG1yPheThrLeuTrYrG1MeLeuTrYrYsProH1sAspLeuG1nProG1yYrYs	645		
Db	481	ACTGATTTACATGTATAGGATCTCTACAAAGCTTCATATCTTACAGCTTGAGAAAGAA	540		
QY	646	TrYrProThrVal1eLeuPhe11eTrYrG1yG1yProG1nValG1nLeuVal1aAsnAsnArgPhe	665		

```
Db      541 TATCCTACTGTGCTGTTCAATATATGATGGTCCACAGTGCAGTTGGTGAATATACGTTT 600
Qy      666 LysGlyValIlyTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValVal 685
Db      601 AAAGGAGTCAGATATTCCGCTTGAAATACCCCTAGCCTCTCTAGATTATGTTGTAGTG 660
Qy      686 IleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyValaPheLysTyrLys 705
Db      661 ATGACACACAGGGGATCCTGTCACCGANGGCTTAATTTGAAGCGCGCTTAATATAA 720
Qy      706 MetGlyGlnIleGluIleAspAspGlnVal 715
Db      721 ATGGGTCAATAGAAATTGACGATCAGTG 750
```

Search completed: October 16, 2003, 03:24:56  
Job time : 5362 secs





DR N-PSDB; AAC85695.  
 XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving  
 PT substrates, identifying inhibitors of DPP8 catalytic activity which  
 PT have therapeutic uses, and for detecting activated T cells -  
 XX  
 XX Claim 5; Page 68-69; 78pp; English.  
 XX  
 CC The sequences given in AAB47188-90 represent fragments of human  
 CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for  
 CC H-Ala-Pro-pNa, H-Gly-Pro-pNa and H-Arg-Pro-pNa. Therefore, it is a  
 CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable  
 CC of hydrolysing the peptide bond C-terminal to Pro in each of these  
 CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for  
 CC cleaving a substrate, and for detecting an activated T cell which  
 CC involves measuring the level of DPP8 gene expression in a T cell. The  
 CC level of DPP8 expression is detected by detecting the amount of DPP8  
 CC RNA in the cell. It is also useful for identifying a molecule capable  
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules  
 CC identified as inhibiting DPP8 catalytic activity may be useful for  
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels  
 CC in non-insulin dependent diabetes mellitus and other disorders  
 CC involving glucose intolerance, enhancing mucosal regeneration and  
 CC as immunosuppressants.  
 CC  
 XX Sequence 310 AA:  
 SQ  
 Query Match 100.0%; Score 1680; DB 22; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-170; Indels 0; Gaps 0;  
 Matches 310; Conservative 0; Mismatches 0;  
 QY 1 FEGRDPSLEHNLVYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60  
 DB 1 FEGRDPSLEHNLVYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60  
 QY 61 YKLSPEDDPCKTKEFPAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 120  
 DB 61 YKLSPEDDPCKTKEFPAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 120  
 QY 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYDIDLDRVGIGHMSYGYSLMALMQR 180  
 DB 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYDIDLDRVGIGHMSYGYSLMALMQR 180  
 QY 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMAQAEKFPSEPNRLLL 240  
 DB 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMAQAEKFPSEPNRLLL 240  
 QY 241 HGFIDENHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300  
 DB 241 HGFIDENHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300  
 QY 301 GSRIALAKVI 310  
 DB 301 GSRIALAKVI 310

RESULT 2  
 AAB08994  
 ID AAB08994 standard; protein; 310 AA.  
 XX  
 AC ABB08994;  
 XX  
 DT 19-JUN-2002 (first entry)  
 XX  
 DE Human dipeptidyl peptidase 8.  
 XX  
 KW Human; dipeptidyl peptidase 8; antidiabetic; antiallergic;  
 KW antiinflammatory.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6337069-B1.  
 XX

PD 08-JAN-2002.  
 XX  
 XX 28-FEB-2001; 2001US-0794236.  
 XX  
 XX 28-FEB-2001; 2001US-0794236.  
 XX  
 PA (BMRA-) BMRA CORP BV.  
 XX  
 PI Grouzmann E, Lacroix J, Monod M;  
 XX  
 DR WPI; 2002-163235/21.  
 XX  
 PT Treating a patient for mucosal inflammation associated with rhinitis,  
 PT sinusitis or both, by intranasally administering a peptidase that  
 PT cleaves at Xaa-Pro sequences, to the patient -  
 XX  
 PS Disclosure; Column 23-24; 13pp; English.  
 XX  
 CC This invention relates to the treating of a patient for mucosal  
 CC inflammation associated with rhinitis or sinusitis, comprising  
 CC intranasally administering a peptidase. The peptidase is considered  
 CC antidiabetic, antiallergic and antiinflammatory in its action.  
 CC The peptidase cleaves at Xaa-Pro sequences and is useful for treating  
 CC a patient for mucosal inflammation associated with rhinitis or  
 CC sinusitis, which is the result of allergies or asthma. This  
 CC sequence represents human dipeptidyl peptidase 8.  
 CC  
 XX Sequence 310 AA:  
 SQ

Query Match 100.0%; Score 1680; DB 23; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-170; Indels 0; Gaps 0;  
 Matches 310; Conservative 0; Mismatches 0;  
 QY 1 FEGRDPSLEHNLVYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60  
 DB 1 FEGRDPSLEHNLVYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60  
 QY 61 YKLSPEDDPCKTKEFPAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 120  
 DB 61 YKLSPEDDPCKTKEFPAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 120  
 QY 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYDIDLDRVGIGHMSYGYSLMALMQR 180  
 DB 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYDIDLDRVGIGHMSYGYSLMALMQR 180  
 QY 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMAQAEKFPSEPNRLLL 240  
 DB 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMAQAEKFPSEPNRLLL 240  
 QY 241 HGFIDENHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300  
 DB 241 HGFIDENHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300  
 QY 301 GSRIALAKVI 310  
 DB 301 GSRIALAKVI 310

RESULT 3  
 AAB47187  
 ID AAB47187 standard; Protein; 882 AA.  
 XX  
 AC AAB47187;  
 XX  
 DT 29-JUN-2001 (first entry)  
 XX  
 DE Human DPP8.  
 XX  
 KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;  
 KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;  
 KW growth hormone deficiency; glucose level; mucosal regeneration;  
 KW non-insulin dependent diabetes mellitus; glucose intolerance;  
 KW immunosuppression.  
 KW

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Active-site /note= "Forms part of Ser-Asp-His catalytic triad"
XX FT Active-site 739
XX FT Active-site 817
XX FT Active-site /note= "Forms part of Ser-Asp-His catalytic triad"
XX FT Active-site 849
XX FT /note= "Forms part of Ser-Asp-His catalytic triad"
XX PN MO200119866-A1.
XX PD 22-MAR-2001.
XX PF 11-SEP-2000; 2000MO-AU01085.
XX PR 10-SEP-1999; 99AU-0002762.
XX PR 18-FEB-2000; 2000AU-0005709.
XX PA (UNSV ) UNIV SYDNEY.
XX PI Abbott CA, Gorell MD;
XX DR WPI; 2001-281520/29.
XX DR N-PSDB; AAC85694.
XX PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
XX PT substrates, identifying inhibitors of DPP8 catalytic activity which
XX PT have therapeutic uses, and for detecting activated T cells
XX PS Claim 1; Fig 2; 78pp; English.
XX CC This sequence represents human dipeptidyl aminopeptidase (DPP8).
XX CC DPP8 has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
XX CC H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
XX CC dipeptidyl peptidase, because it is capable of hydrolysing the
XX CC peptide bond C-terminal to Pro in each of these compounds. DPP8
XX CC is homologous with human DPPIV. DPP8 is useful for cleaving a
XX CC substrate, and for detecting an activated T cell which involves
XX CC measuring the level of DPP8 gene expression in a T cell. The level
XX CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
XX CC in the cell. It is also useful for identifying a molecule capable
XX CC of inhibiting the cleavage of the substrate by DPP8. Molecules
XX CC identified as inhibiting DPP8 catalytic activity may be useful for
XX CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
XX CC in non-insulin dependent diabetes mellitus and other disorders
XX CC involving glucose intolerance, enhancing mucosal regeneration and
XX CC as immunosuppressants.
XX SQ Sequence 882 AA;

Query Match 97.9%; Score 1645.5; DB 22; Length 882;
Best Local Similarity 86.4%; Pred. No. 1.9e-165;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 1 FEETKDSPLLEHLLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISRYSNQKNPHCVSL 60
DB 524 FEETKDSPLLEHLLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISRYSNQKNPHCVSL 583
QY 61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKPHDLQPG 120
DB 584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKPHDLQPG 643
QY 121 KKPTVLFITGGPQ----- 134
DB 644 KKPTVLFITGGPQVQVQVNNRFGKVYFRLNTLASLGVVVVVINDRGSCHRGKLFEGAFK 703
QY 135 ---GOEIDDOVEGLQYTLASRYDFIDLRVGIHGMSYGYSLMALMQRSDFRVAIAGA 191
DB 704 YKMGQIETDOVEGLQYTLASRYDFIDLRVGIHGMSYGYSLMALMQRSDFRVAIAGA 763
QY 192 PVTLMIFYDTGYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFIDENVHFA 251

```

```

DB 764 PVTLMIFYDTGYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFIDENVHFA 823
QY 252 HTSILSLFVNAAGRPYDQIYPOERHSIRVSESGEHLHLHTYLOENLGSRIALKYI 310
DB 824 HTSILSLFVNAAGRPYDQIYPOERHSIRVSESGEHLHLHTYLOENLGSRIALKYI 882

RESULT 4
ID AAE24170
ID AAE24170 standard; Protein, 882 AA.
XX AC AAE24170;
XX AC AAE24170;
XX DT 23-SEP-2002 (first entry)
XX DE Human dipeptidyl peptidase 8 (DPP8) protein.
XX KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
XX KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
XX KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
XX KW antiviral; enzyme.
XX OS Homo sapiens.
XX AC MO200234900-A1.
XX AC MO200234900-A1.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001MO-AU01388.
XX PR 27-OCT-2000; 2000AU-0001078.
XX PA (UNSV ) UNIV SYDNEY.
XX PI Abbott CA, Gorell MD;
XX DR WPI; 2002-454646/48.
XX DR N-PSDB; AAD36956.
XX PT New dipeptidyl peptidase (DPP) peptidase, useful for screening
XX PT inhibitors of DPP catalytic activity, which may be employed to treat
XX PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX PT rejection and HIV infection -
XX PS Example; Fig 1; 91pp; English.
XX CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX CC polynucleotides encoding such proteins. The DPP peptidase are useful for
XX CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX CC rejection and HIV (human immuno deficiency virus) infection. The present
XX CC sequence is human DPP8 protein.
XX SQ Sequence 882 AA;

Query Match 97.9%; Score 1645.5; DB 23; Length 882;
Best Local Similarity 86.4%; Pred. No. 1.9e-165;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 1 FEETKDSPLLEHLLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISRYSNQKNPHCVSL 60
DB 524 FEETKDSPLLEHLLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISRYSNQKNPHCVSL 583
QY 61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKPHDLQPG 120
DB 584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKPHDLQPG 643
QY 121 KKPTVLFITGGPQ----- 134
DB 644 KKPTVLFITGGPQVQVQVNNRFGKVYFRLNTLASLGVVVVVINDRGSCHRGKLFEGAFK 703
QY 135 ---GOEIDDOVEGLQYTLASRYDFIDLRVGIHGMSYGYSLMALMQRSDFRVAIAGA.191

```

Db 704 YKMQIETIDDOVEGLQYLASRYDFIDLRVGIHGMSCYGYLSLWALMQRSDIFRVAIAGA 763  
 Qy 192 PVTLMIFDYGTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLIHGFLENVFA 251  
 Db 764 PVTLMIFDYGTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLIHGFLENVFA 823  
 Qy 252 HTSILSFLVAGKPYDLOIYPOERHSIRVPSGEHVELHLHYLOENLGSRIALAKVI 310  
 Db 824 HTSILSFLVAGKPYDLOIYPOERHSIRVPSGEHVELHLHYLOENLGSRIALAKVI 882

RESULT 5  
 ABG61591  
 ID ABG61591 standard; Protein; 882 AA.  
 AC ABG61591;  
 XX  
 XX 12-AUG-2002 (first entry)  
 DT  
 DE Human DPPIV related serine protease DPP-1.  
 XX  
 KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;  
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KM dyskinesia; reproductive disorder; inflammatory disorder;  
 KM metabolic disorder.  
 KM  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200231134-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 12-CT-2001; 2001WO-US31874.  
 XX  
 PR 12-OCT-2000; 2000US-240117P.  
 XX  
 PA (FERR) FERRING BV.  
 XX  
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;  
 XX  
 DR WPI; 2002-444178/47.  
 DR N-PSDB; ABK83322.  
 XX  
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and  
 PT viral infections, cancers, allergies, neurological disorders, or pain  
 PT  
 PS Claim 17; Fig 1; 113p; English.  
 XX  
 XX The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related  
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABG61591-ABG61612 represent human DPP proteins.  
 XX  
 SQ Sequence 882 AA;  
 Query Match 97.9%; Score 1645.5; DB 23; Length 882;  
 Best Local Similarity 86.4%; Pred. No. 1.9e-165;

Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;  
 Qy 1 FEETKDSPLREHLVYVYVNGEYTRLTDRGYSHSCCISOHCDPFISSYNOKNPHCVSL 60  
 Db 524 FEETKDSPLREHLVYVYVNGEYTRLTDRGYSHSCCISOHCDPFISSYNOKNPHCVSL 583  
 Qy 61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEISFSESTTGFTLYGMLYKPHDLOG 120  
 Db 584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEISFSESTTGFTLYGMLYKPHDLOG 643  
 Qy 121 KKYPTVLFYGGPQ----- 134  
 Db 644 KKYPTVLFYGGPQVOLLVNNFKGVKYYRLNTLASLGIVVVINDRSGCHRLKFECAFK 703  
 Qy 135 --GQIETIDDOVEGLQYLASRYDFIDLRVGIHGMSCYGYLSLWALMQRSDIFRVAIAGA 191  
 Db 704 YKMQIETIDDOVEGLQYLASRYDFIDLRVGIHGMSCYGYLSLWALMQRSDIFRVAIAGA 763  
 Qy 192 PVTLMIFDYGTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLIHGFLENVFA 251  
 Db 764 PVTLMIFDYGTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLIHGFLENVFA 823  
 Qy 252 HTSILSFLVAGKPYDLOIYPOERHSIRVPSGEHVELHLHYLOENLGSRIALAKVI 310  
 Db 824 HTSILSFLVAGKPYDLOIYPOERHSIRVPSGEHVELHLHYLOENLGSRIALAKVI 882

RESULT 6  
 AAU74749  
 ID AAU74749 standard; Protein; 882 AA.  
 AC AAU74749;  
 XX  
 XX 09-APR-2002 (first entry)  
 DT  
 DE Human protease PRTS-9 protein sequence.  
 XX  
 KM Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;  
 KM cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;  
 KM inflammatory disorder; acquired immunodeficiency syndrome; AIDS;  
 KM cell proliferative disorder; developmental disorder; epilepsy;  
 KM Duchenne muscular dystrophy; epithelial disorder; neurological disorder;  
 KM reproductive disorder; endometriosis.  
 KM  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200198468-A2.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 13-JUN-2001; 2001WO-US19178.  
 XX  
 PR 16-JUN-2000; 2000US-212336P.  
 PR 22-JUN-2000; 2000US-213955P.  
 PR 29-JUN-2000; 2000US-215396P.  
 PR 07-JUL-2000; 2000US-216821P.  
 PR 14-JUL-2000; 2000US-218946P.  
 XX  
 PA (INCYTE) INCYTE GENOMICS INC.  
 XX  
 PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM,  
 PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA,  
 PI Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT,  
 PI Azimzal Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L,  
 PI Kallik DA;  
 XX  
 DR WPI; 2002-090437/12.  
 DR N-PSDB; ABK12892.  
 XX  
 PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful  
 PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.  
 PT gastritis), cardiovascular (e.g. atherosclerosis) and cell  
 PT proliferative (e.g. cancer) disorders -

XX Claim 1; Page 140-142; 177pp; English.

PS The present invention relates to twenty one new human proteases,

XX referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and

CC polypeptides of the invention are useful in the diagnosis, treatment and

CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and

CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and

CC myocardial infarction, autoimmune/inflammatory e.g. acquired

CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell

CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker

CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.

CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and

CC endometriosis disorders. Numerous other examples of each disorder are

CC given in the specification. The present protein sequence represents

CC the human protease PRTS-9 protein of the invention.

XX

XX Sequence 882 AA;

XX

Query Match 97.9%; Score 1645.5; DB 23; Length 882;

Best Local Similarity 86.4%; Pred. No. 1.9e-165;

Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 1 FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 60

Db FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 583

QY 524 FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 583

Db 524 FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 583

QY 61 YKLSPPEDPTCKKEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLQPG 120

Db 584 YKLSPPEDPTCKKEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLQPG 643

QY 121 KKYPVTLFIYGGPQ----- 134

Db 644 KKYPVTLFIYGGPQVQVQVNNRFGKVKYFRNLNTASLGVVVVIDNRGSHGKKEGAFK 703

QY 135 ---GQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRYAIGA 191

Db 704 YKMQCQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRYAIGA 763

QY 192 PVTLMIFPDGTYTERYMGHPDQNEGGYLGSAVMAQKFPSEPNRLLLHGFLENVHFA 251

Db 764 PVTLMIFPDGTYTERYMGHPDQNEGGYLGSAVMAQKFPSEPNRLLLHGFLENVHFA 823

QY 252 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESEGEHLLHLHYQENLGSRIALAKYI 310

Db 824 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESEGEHLLHLHYQENLGSRIALAKYI 882

RESULT 7

AAAG78415

XX AAG78415 standard; Protein: 882 AA.

XX

XX AAG78415;

XX

XX 12-APR-2002 (first entry)

XX

XX Amino acid sequence of 21953 human prolyl oligopeptidase.

XX

XX 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;

XX cancer; cardiovascular disease; autoimmune disease; atopic allergy;

XX neuronal disorder; vascular disorder; prostate disorder; cystostatic;

XX antidiabetic; antiautistic; antiautismic; antiinflammatory;

XX diabetes mellitus; arthritis; multiple sclerosis; asthma;

XX Grave's disease; neuronal disorder; demyelinating disease.

XX

XX Homo sapiens.

XX

XX PN MO200179473-A2.

XX

XX PD 25-OCT-2001.

XX

XX PF 11-APR-2001; 2001WO-US40483.

XX

PR 18-APR-2000; 2000US-197508P.

XX

XX (MILL-) MILLENNIUM PHARM INC.

XX

XX Meyers RA, Williamson M;

XX

XX WPI; 2002-03435/04.

XX

XX N-PsDB; AAH99934.

XX

XX New polypeptides 21953, member of human prolyl oligopeptidase family,

XX useful as diagnostic targets and therapeutic agents for controlling

XX cancer, lymphoma and leukemia

XX

XX Claim 1; Page 102-103; 121pp; English.

XX

XX This invention relates to an isolated 21953 human prolyl

XX oligopeptidase. Which is cytosolic, antidiabetic, antiautistic,

XX neuroprotective, antihypertensive, dermatological, antipsoriatic,

XX antiautismic, ophthalmological, antiinflammatory, nootropic,

XX antiparkinsonian, anticonvulsant, gynaecological, vasotropic,

XX antitumor, cardiac, antiatherosclerotic, anorectic and

XX metabolic in its action. Uses include gene therapy, expression or

XX activity of 21953 protein modulator, it is useful for identifying a

XX compound which binds to it and can be used in preventing, treating

XX or detecting a cellular proliferative or differentiative disorder.

XX The 21953 molecules can act as novel diagnostic targets and therapeutic

XX agents for controlling disorders associated with the aberrant activity

XX or degradation of peptide hormones e.g., disorders associated with cell

XX differentiation and proliferation such as cancer, immune function,

XX reproductive, neurological and cardiovascular function. The 21953

XX molecules are thus useful for treating and preventing cellular

XX proliferative and differentiative disorders, hematopoietic neoplastic

XX disorders, immune disorders such as autoimmune diseases, diabetes

XX mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,

XX neuronal disorders, demyelinating diseases, vascular disorders and

XX metabolism or pain disorders. This sequence represents the amino

XX acid sequence of 21953 human prolyl oligopeptidase.

XX

XX Sequence 882 AA;

XX

XX

Query Match 97.9%; Score 1645.5; DB 23; Length 882;

Best Local Similarity 86.4%; Pred. No. 1.9e-165;

Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 1 FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 60

Db FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 583

QY 524 FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 583

Db 524 FEETKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHCCTISQHCDFRISKYSNQKNHCVSL 583

QY 61 YKLSPPEDPTCKKEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLQPG 120

Db 584 YKLSPPEDPTCKKEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLQPG 643

QY 121 KKYPVTLFIYGGPQ----- 134

Db 644 KKYPVTLFIYGGPQVQVQVNNRFGKVKYFRNLNTASLGVVVVIDNRGSHGKKEGAFK 703

QY 135 ---GQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRYAIGA 191

Db 704 YKMQCQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRYAIGA 763

QY 192 PVTLMIFPDGTYTERYMGHPDQNEGGYLGSAVMAQKFPSEPNRLLLHGFLENVHFA 251

Db 764 PVTLMIFPDGTYTERYMGHPDQNEGGYLGSAVMAQKFPSEPNRLLLHGFLENVHFA 823

QY 252 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESEGEHLLHLHYQENLGSRIALAKYI 310

Db 824 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESEGEHLLHLHYQENLGSRIALAKYI 882

RESULT 8

ABU07720

XX ABU07720 standard; Protein: 882 AA.

XX

AC ABU07720;  
 XX  
 DT 19-MAY-2003 (first entry)  
 XX  
 DE Human serine protease HIPHUM46.  
 XX  
 KW Human; enzyme; HIPHUM46; serine protease; gene therapy; osteoarthritis;  
 KW serine protease activity modulation; dipeptidyl peptidase activity;  
 KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;  
 KW Alzheimer's disease; paraneoplastic palsy; Huntington's disease;  
 KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;  
 KW irritable bowel syndrome; type I diabetes; fecal incontinence; anaemia;  
 KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;  
 KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;  
 KW multiple sclerosis.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 259..260  
 FT /note="Paired glutamates of the beta propeller domain"  
 FT Active-site 739  
 FT /label= Catalytic\_serine\_residue  
 FT Active-site 817  
 FT /label= Catalytic\_aspartate\_residue  
 FT Active-site 849  
 FT /label= Catalytic\_histidine\_residue  
 XX  
 PN GB2374869-A.  
 XX  
 PD 30-OCT-2002.  
 XX  
 PF 22-JAN-2002; 2002GB-0001404.  
 XX  
 PR 23-JAN-2001; 2001GB-0001760.  
 XX  
 PA (GLAXO ) GLAXO GROUP LTD.  
 PI Edbrooke MR, Lewis AP;  
 XX  
 DR MPI; 2003-150703/15.  
 DR N-PSDB; ABX12255.  
 XX  
 PT Identifying modulators of serine protease activity useful for treating  
 PT musculoskeletal diseases, by contacting cell expressing a novel serine  
 PT protease polypeptide with a compound and monitoring serine protease  
 PT activity -  
 XX  
 PS Claim 10; Page 26-29; 38pp; English.  
 XX  
 CC The invention relates to a method of identifying a substance that  
 CC modulates serine protease activity, comprising contacting a cell such as  
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a  
 CC virus, expressing a serine protease polypeptide (HIPHUM 46), or its  
 CC variant having dipeptidyl peptidase activity, or a serine protease  
 CC isolated from the cell with a test substance and monitoring for serine  
 CC protease activity. The method is useful for identifying a substance that  
 CC modulates serine protease activity. A modulator of the serine protease is  
 CC useful in the manufacture of a medicament for treatment or prophylaxis of  
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus  
 CC infection, Alzheimer's disease, paraneoplastic palsy, myotonic  
 CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.  
 CC Additional disease that may be treated using modulators of the serine  
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung  
 CC disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,  
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,  
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple  
 CC sclerosis. The present sequence represents the amino acid sequence of the  
 CC human serine protease HIPHUM46.  
 XX  
 SQ Sequence 882 AA;  
 Query Match 97.9%; Score 1645.5; DB 24; Length 882;

Best Local Similarity 86.4%; Pred. No. 1,9e-165;  
 Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;  
 QY 1 FEGTGDSPLEHLLVYVSYVNPGEVTRLTDRGYSHSCCISQHCDFISKYSNQNPHCVSL 60  
 DB 524 FEGTGDSPLEHLLVYVSYVNPGEVTRLTDRGYSHSCCISQHCDFISKYSNQNPHCVSL 583  
 QY 61 YKLSPEDDPTCKTKEFMATILDSAGPLDPYRPEIFSESTTGFTLYGMLYKPHDLQFG 120  
 DB 584 YKLSPEDDPTCKTKEFMATILDSAGPLDPYRPEIFSESTTGFTLYGMLYKPHDLQFG 643  
 QY 121 KKPPTVLFYGGPQ----- 134  
 DB 644 KKPPTVLFYGGPQVQLVNNRFGYKFRLLNTLASIGYVVVVIDNRGSGRGLKFECAFK 703  
 QY 135 ---GQIEIDDOVEGLOYLASRYDFIDLDVRGIGHGMSYGGYLSLMALMQRSDIFRVAIAGA 191  
 DB 704 YKMGQIEIDDOVEGLOYLASRYDFIDLDVRGIGHGMSYGGYLSLMALMQRSDIFRVAIAGA 763  
 QY 192 PVTLMIFDYDTGYTERYMGHPDNEQGYLYGSVAMQAEKPPSEPNRLLLHGFLENVHFA 251  
 DB 764 PVTLMIFDYDTGYTERYMGHPDNEQGYLYGSVAMQAEKPPSEPNRLLLHGFLENVHFA 823  
 QY 252 HTSILSFLVRAGKPYDQIYPOERHSIRVPESGEHYELHLHYLOENIGSRITAAIKVI 310  
 DB 824 HTSILSFLVRAGKPYDQIYPOERHSIRVPESGEHYELHLHYLOENIGSRITAAIKVI 882  
 RESULT 9  
 AAB93565  
 ID AAB93565 standard; Protein; 632 AA.  
 XX  
 AC AAB93565;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:12964.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-)%HELIX RES INGT.  
 XX  
 PI Oca T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J,  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR MPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 12964; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 632 AA;

Query Match 83.4%; Score 1401; DB 22; Length 632;  
 Best Local Similarity 84.6%; Pred. No. 1.2e-139;  
 Matches 269; Conservative 11; Mismatches 20; Indels 18; Gaps 5;

QY 1 FEGTDSPLEHLLYVSVYNPEGVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 60  
 DB 325 FEGTDSPLEHLLYVSVYNPEGVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 384  
 QY 61 YKLSSPEDDPCTCKTEFWATITLDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQ 120  
 DB 385 YKLSSPEDDPCTCKTEFWATITLDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQ 444  
 QY 121 KKYPTVLFYGGPQGOIEIDQVEGLQY----LAS-RYDFIDLDRVGI--HGWSYGYL 172  
 DB 445 KKYPTVLFYGGPQGOIEIDQVEGLQY----LAS-RYDFIDLDRVGI--HGWSYGYL 503  
 QY 173 SLMALMQRSDIFRVAIAGAPVTLMIFYDTGTERYWGHPDQNEGGYLLGSVAMQAEKPPS 232  
 DB 504 KYK-----MVAIAGAPVTLMIFYDTGTERYWGHPDQNEGGYLLGSVAMQAEKPPS 554  
 QY 233 EPNRLLLHGFLENVHFAHTSILSLFVRAGKPYDLOIYPOERHSIRVPESGEHYELHL 292  
 DB 555 EPNRLLLHGFLENVHFAHTSILSLFVRAGKPYDLOIYPOERHSIRVPESGEHYELHL 614  
 QY 293 LHYLOENLGSRIALAKVI 310  
 DB 615 LHYLOENLGSRIALAKVI 632

RESULT 10

AAAB47189 standard; Protein; 465 AA.

AAAB47189;

29-JUN-2001 (first entry)

Human DPP8 318Thr-658Val+759Ala-882Ile.

Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;  
 dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;  
 growth hormone deficiency; glucose level; mucosal regeneration;  
 non-insulin dependent diabetes mellitus; glucose intolerance;  
 immunosuppression.

Homo sapiens.

MO200119866-A1.

22-MAR-2001.

11-SEP-2000; 2000MO-AU01085.

XX

PR 10-SEP-1999; 99AU-0002762.  
 PR 18-FEB-2000; 2000AU-0005709.  
 XX (UNSY ) UNIV SYDNEY.  
 PA Abbott CA, Gorell MD;  
 PI WPI; 2001-281520/29.  
 DR N-PSDB; AAC65696.  
 XX  
 PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving  
 PT substrates, identifying inhibitors of DPP8 catalytic activity which  
 PT have therapeutic uses, and for detecting activated T cells  
 PS Claim 5; Page 71-72; 78pp; English.  
 XX  
 CC The sequences given in AAB47188-90 represent fragments of human  
 CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for  
 CC H-Ala-Pro-PNA, H-Gly-Pro-PNA and H-Arg-Pro-PNA. Therefore, it is a  
 CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable  
 CC of hydrolysing the peptide bond C-terminal to Pro in each of these  
 CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for  
 CC cleaving a substrate, and for detecting an activated T cell which  
 CC involves measuring the level of DPP8 gene expression in a T cell. The  
 CC level of DPP8 expression is detected by detecting the amount of DPP8  
 CC RNA in the cell. It is also useful for identifying a molecule capable  
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules  
 CC identified as inhibiting DPP8 catalytic activity may be useful for  
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels  
 CC in non-insulin dependent diabetes mellitus and other disorders  
 CC involving glucose intolerance, enhancing mucosal regeneration and  
 CC as immunosuppressants.  
 XX  
 SQ Sequence 465 AA;  
 Query Match 82.0%; Score 1377.5; DB 22; Length 465;  
 Best Local Similarity 83.5%; Pred. No. 2.4e-137;  
 Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 FEGTDSPLEHLLYVSVYNPEGVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 60  
 DB 207 FEGTDSPLEHLLYVSVYNPEGVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 266  
 QY 61 YKLSSPEDDPCTCKTEFWATITLDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQ 120  
 DB 267 YKLSSPEDDPCTCKTEFWATITLDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQ 326  
 QY 121 KKYPTVLFYGGPQGOIEIDQVEGLQYLAIRYDFIDLDRVGIHGWSYGYLSMALMQR 180  
 DB 327 KKYPTVLFYGGPQGOIEIDQVEGLQYLAIRYDFIDLDRVGIHGWSYGYLSMALMQR 340  
 QY 181 SDIFRVAIAGAPVTLMIFYDTGTERYWGHPDQNEGGYLLGSVAMQAEKPPSEPNRLLL 240  
 DB 341 -----VAIAGAPVTLMIFYDTGTERYWGHPDQNEGGYLLGSVAMQAEKPPSEPNRLLL 395  
 QY 241 HGFLENVHFAHTSILSLFVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENL 300  
 DB 396 HGFLENVHFAHTSILSLFVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENL 455  
 QY 301 GSRIALAKVI 310  
 DB 456 GSRIALAKVI 465

RESULT 11

ABAB97362 standard; Protein; 724 AA.

ABAB97362;

27-JUN-2002 (first entry)

Novel human protein SEQ ID NO: 630.

DE

```
XX Human; anti-naemic; vulnerary; anti-inflammatory; immunomodulator;
XX anti-infectivity; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX neuroprotective; anti-parkinsonian; protein therapy; EST;
XX expressed sequence tag.
OS Homo sapiens.
XX
XX WO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US26015.
XX
XX 11-SEP-2000; 2000US-0659671.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX N-PSDB; ABN32548.
XX
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX
XX Example 2; SEQ ID NO 630; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions
XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention.
XX
XX Sequence 724 AA:
XX
XX Query Match 82.0%; Score 1377.5; DB 23; Length 724;
XX Best Local Similarity 83.5%; Pred. No. 4.6e-137;
XX Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;
XX
XX 1 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDFISKYSNOKNPHCVSL 60
XX 466 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDFISKYSNOKNPHCVSL 525
XX
XX 61 YKLSSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPG 120
XX 526 YKLSSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPG 585
XX
XX 121 KKYPTVLFYGGPQGIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 180
XX 566 KKYPTVLFYGGPQGIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 599
XX
XX 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 240
XX 600 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 654
XX
XX 241 HGFIDENVFAHTSILLSFVRAGKPYDLOIYPOERHSIRVSESGEYELHLHYLOENTL 300
XX 655 HGFIDENVFAHTSILLSFVRAGKPYDLOIYPOERHSIRVSESGEYELHLHYLOENTL 714
XX
XX 301 GSRIATLAKVI 310
XX 715 GSRIATLAKVI 724
XX
XX RESULT 12
XX ABB97361
```

```
ID ABB97361 standard; Protein; 782 AA.
XX
XX ABB97361;
XX
XX 27-JUN-2002 (first entry)
XX
XX Novel human protein SEQ ID NO: 629.
XX
XX Human; anti-naemic; vulnerary; anti-inflammatory; immunomodulator;
XX anti-infectivity; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX neuroprotective; anti-parkinsonian; protein therapy; EST;
XX expressed sequence tag.
XX
XX Homo sapiens.
XX
XX WO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US26015.
XX
XX 11-SEP-2000; 2000US-0659671.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX N-PSDB; ABN32547.
XX
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX
XX Example 2; SEQ ID NO 629; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions
XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention.
XX
XX Sequence 782 AA:
XX
XX Query Match 82.0%; Score 1377.5; DB 23; Length 782;
XX Best Local Similarity 83.5%; Pred. No. 5.1e-137;
XX Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;
XX
XX 1 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDFISKYSNOKNPHCVSL 60
XX 524 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDFISKYSNOKNPHCVSL 583
XX
XX 61 YKLSSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPG 120
XX 584 YKLSSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPG 643
XX
XX 121 KKYPTVLFYGGPQGIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 180
XX 644 KKYPTVLFYGGPQGIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 657
XX
XX 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 240
XX 658 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 712
XX
XX 241 HGFIDENVFAHTSILLSFVRAGKPYDLOIYPOERHSIRVSESGEYELHLHYLOENTL 300
XX 713 HGFIDENVFAHTSILLSFVRAGKPYDLOIYPOERHSIRVSESGEYELHLHYLOENTL 772
XX
XX Db
```

QY 301 GSRIALAKVI 310  
 DB 773 GSRIALAKVI 782

RESULT 13  
 ID AAE14337 standard; Protein; 580 AA.  
 AC AAE14337;  
 XX  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Human protease PRTS-2 protein.  
 XX  
 KM Human: protease; PRTS-2; tranquiliser; gene therapy; vaccine; allergy;  
 KM infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
 KM gastroenteritis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
 KM gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
 KM epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;  
 KM hypertension; neurological disorder; Parkinson's disease; drug screening;  
 KM cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;  
 KM diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;  
 KM autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
 KM developmental disorder; reproductive disorder; infertility; diarrhoea;  
 KM dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice.  
 OS Homo sapiens.  
 XX  
 PN WO200183775-A2.  
 XX  
 PD 08-NOV-2001.  
 XX  
 PF 04-MAY-2001; 2001WO-US14651.  
 XX  
 PR 04-MAY-2000; 2000US-202082P.  
 PR 11-MAY-2000; 2000US-203566P.  
 PR 17-MAY-2000; 2000US-205803P.  
 PR 25-MAY-2000; 2000US-207477P.  
 PR 01-JUN-2000; 2000US-209402P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Deleageane AM, Lal P, Hafalia A, Patterson C, Wallia NK, Kearney L,  
 PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;  
 PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DAM;  
 PI Reddy R, Yue H, Tang YT;  
 XX  
 DR WPI; 2002-034518/04.  
 DR N-PSDB; AAD23843.  
 XX  
 PT Novel human proteases and polynucleotides encoding the proteases,  
 PT useful for treating, diagnosing or preventing cell proliferative,  
 PT cardiovascular, autoimmune/inflammatory, neurological and developmental  
 PT disorders -  
 XX  
 PS Claim 1; Page 120-121; 151pp; English.  
 XX  
 CC The invention relates to human proteases (PRTS1-14) and its corresponding  
 CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for  
 CC the diagnosis, treatment and prevention of disorders associated with  
 CC increased or decreased expression of PRTS. Examples of such disorders  
 CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
 CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders  
 CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,  
 CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
 CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
 CC viral, bacterial, fungal, parasitic, protozoal and helminthic  
 CC infections); cardiovascular disorders (myocardial infarction, ischaemic  
 CC heart disease and hypertension); neurological disorders (epilepsy,  
 CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,  
 CC Parkinson's disease, stroke, mental disorders including mood, anxiety  
 CC and seasonal affective disorder and prion diseases); gastrointestinal

CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);  
 CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,  
 CC alopecia, scabies, insect bites and urticaria); reproductive disorder  
 CC (infertility, disruption of estrous and menstrual cycle and  
 CC gynaecomastia); and developmental disorders (renal tubular acidosis,  
 CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
 CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic  
 CC fragments are useful for screening libraries of compounds in several drug  
 CC screening assays. The present sequence is human protease PRTS-2 protein.  
 XX

Sequence 580 AA:  
 Query Match 81.7%; Score 1372; DB 23; Length 580;  
 Best Local Similarity 83.6%; Pred. No. 1,3e-136;  
 Matches 266; Conservative 11; Mismatches 19; Indels 22; Gaps 6;

QY 1 FECKTSDPLEHLLVYVSYVNPBEVTRLTDGRVSHSCCISQHCDFISKYSNQNPHCVSL 60  
 DB 277 FECKTSDPLEHLLVYVSYVNPBEVTRLTDGRVSHSCCISQHCDFISKYSNQNPHCVSL 336  
 QY 61 YKLSPEDDPTCKTKEFMAATILDSAGPLDYPTRPEIFSFESTTGFTLYGMLYKPHLOPG 120  
 DB 337 YKLSPEDDPTCKTKEFMAATILDSAGPLDYPTRPEIFSFESTTGFTLYGMLYKPHLOPG 396  
 QY 121 KKYPTVLFYGGPGQGEIDDOVEGLQY-----LAS-RDFIDLDKRVGI--HGMVGYGL 172  
 DB 397 KKYPTVLFYGGPGQVL-VNMRFKGVKFRMLTSLAGLYVVVYIDNRGSGHRLKEGAF 455  
 QY 173 SLMALMQRSDIFRVAIAGAPVTLMIFDYGTERVYVGHDPQNEQGYLLGSVMAQAEKFP 232  
 DB 456 KYK-----MVAIAGAPVTLMIFDYGTERVYVGHDPQNEQGYLLGSVMAQAEKFP 506  
 QY 233 EPNRLILHGFIDENVHFAHTSILSLFLVRACKPDLQIYPERHRIKPESEHEHL 292  
 DB 507 EPNRLILHGFIDENVHFAHTSILSLFLVRACKPDLQIYPERHRIKPESEHEHL 562  
 QY 293 LHYIQENIGSRIALAKVI 310  
 DB 563 LHYIQENIGSRIALAKVI 580

RESULT 14  
 ID AAE24169 standard; Protein; 869 AA.  
 XX  
 AC AAE24169;  
 XX  
 DT 23-SEP-2002 (first entry)  
 XX  
 DE Alternative version of murine dipeptidyl peptidase 9 (DPP9) protein.  
 XX  
 KM Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;  
 KM human immunodeficiency virus; graft rejection; cytostatic; autoimmunity;  
 KM type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;  
 KM antiviral; enzyme.  
 XX  
 OS Mus sp.  
 XX  
 PN WO200234900-A1.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-AU01388.  
 XX  
 PR 27-OCT-2000; 2000AU-0001078.  
 PR (UNSY) UNIV SYDNEY.  
 PA  
 XX  
 PI Abbott CA, Gorrall MD;  
 XX  
 DR WPI; 2002-454646/48.  
 DR N-PSDB; AAD38955.  
 XX



PT New dipeptidyl peptidase (DPP) peptidases, useful for screening  
 PT inhibitors of DPP catalytic activity, which may be employed to treat  
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft  
 PT rejection and HIV infection -  
 XX  
 PS Claim 1; Page 70-74; 91pp; English.  
 XX  
 CC The present invention relates to dipeptidyl peptidase (DPP) proteins and  
 CC polynucleotides encoding such proteins. The DPP peptidases are useful for  
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful  
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft  
 CC rejection and HIV (human immuno deficiency virus) infection. The present  
 CC sequence is an alternative version of murine DPP9 protein.  
 CC Note: This sequence is stated to be the same as that shown as  
 CC SEQ ID NO: 4 in the sequence listing of the specification. However these  
 CC sequences differ.  
 XX  
 SQ Sequence 869 AA;  
 XX  
 Query Match 65.1%; Score 1094.5; DB 23; Length 869;  
 Best Local Similarity 57.3%; Pred. No. 7.5e-107;  
 Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;  
 QY 1 FEGTKSPLEHNLVYVSYNPGEVRLTRDRGYSHSCCISQHCDFPISKYSNQNPHCVSL 60  
 DB 521 FQGTNDTPLEHNLVYVSYNPGEVRLTRDRGYSHSCCISQHCDFPISKYSNQNPHCVSV 580  
 QY 61 YKLSSPEDPTCKTEKFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQ 120  
 DB 581 YKLSGDDDDPLKHQKPRFMAWMMAANCPDYVPELFFHTRADVOLYGMITKPHALDPQ 640  
 QY 121 KKYPTVLFYGGPO----- 134  
 DB 641 RGHPTVLFYGGPOVOLVNNSEFKGIKYLRLNTLASLGYAVVVIDGRGSCORGLRFEGLAK 700  
 QY 135 ---GQIEIDQVGEGLQYASRYDFIDLRVGHGMSYGGYLSLMALMORSDFRVAIAGA 191  
 DB 701 NQMGQVEIDQVGEGLQYVAEKYGFIDLSRAVHGMVYGGYLSLMGLIHKQVFKVAIAGA 760  
 QY 192 PVTLMIFDYTGTYTERMGHPDONEOGYLGSVAMQAEKFPSEPNRLLILHGFLENVHFA 251  
 DB 761 PVTVMMAVDTGYTERMDVPENNNOGYEAGSVALLHVEKLPNEPNRLLILHGFLENVHVF 820  
 QY 252 HTSILSFLVRACKPYDLOIYQERHSIRVPSGSEHYEHLHLHYQENL 300  
 DB 821 HTNPLVSQILIRACKPYOLOIYFNERHSIRCPESGEHYEVLTLHFLQEHL 869

RESULT 15  
 ABG64845  
 ID ABG64845 standard; Protein; 497 AA.  
 XX  
 AC ABG64845;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE Human albumin fusion protein #1520.  
 XX  
 KW Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytotoxic; antifertility; antinflammatory; anticancer;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 OS  
 XX WO20017137-A1.  
 XX  
 PD 18-OCT-2001.

XX  
 XX 12-APR-2001; 2001WO-US11988.  
 PF  
 XX 12-APR-2000; 2000US-229358P.  
 PR 25-APR-2000; 2000US-199384P.  
 PR 21-DEC-2000; 2000US-256931P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Haeltline WA;  
 DR WPI; 2002-010886/01.  
 XX  
 PT New fusion protein for treating disease e.g. diabetes comprises an  
 PT albumin fused to a therapeutic protein -  
 XX  
 PS Claim 1; Page 1544-1545; 2102pp; English.  
 XX  
 CC The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA), also known as human serum  
 CC albumin (HSA). The proteins are useful for treating a disease or  
 CC disorder that may be modulated by therapeutic protein X. The albumin  
 CC extends the shelf-life of protein X, and may increase its biological  
 CC in vitro/in vivo activity. The protein is useful for creating and  
 CC diagnosing disorders such as cancer, reproductive disorders, digestive  
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders  
 CC (e.g. diabetes), haematopoietic disorders, neural disorders  
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders  
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin  
 CC fusion proteins of the invention.  
 XX  
 SQ Sequence 497 AA;  
 XX  
 Query Match 65.1%; Score 1093.5; DB 23; Length 497;  
 Best Local Similarity 57.3%; Pred. No. 4.2e-107;  
 Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;  
 QY 1 FEGTKSPLEHNLVYVSYNPGEVRLTRDRGYSHSCCISQHCDFPISKYSNQNPHCVSL 60  
 DB 149 FQGTNDTPLEHNLVYVSYNPGEVRLTRDRGYSHSCCISQHCDFPISKYSNQNPHCVSV 208  
 QY 61 YKLSSPEDPTCKTEKFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQ 120  
 DB 209 YKLSGDDDDPLKHQKPRFMAWMMAANCPDYVPELFFHTRSDVRLYGMITKPHALDPQ 268  
 QY 121 KKYPTVLFYGGPO----- 134  
 DB 269 KGHPTVLFYGGPOVOLVNNSEFKGIKYLRLNTLASLGYAVVVIDGRGSCORGLRFEGLAK 328  
 QY 135 ---GQIEIDQVGEGLQYASRYDFIDLRVGHGMSYGGYLSLMALMORSDFRVAIAGA 191  
 DB 329 NQMGQVEIDQVGEGLQYVAEKYGFIDLSRAVHGMVYGGYLSLMGLIHKQVFKVAIAGA 388  
 QY 192 PVTLMIFDYTGTYTERMGHPDONEOGYLGSVAMQAEKFPSEPNRLLILHGFLENVHFA 251  
 DB 389 PVTVMMAVDTGYTERMDVPENNNOGYEAGSVALLHVEKLPNEPNRLLILHGFLENVHVF 448  
 QY 252 HTSILSFLVRACKPYDLOIYQERHSIRVPSGSEHYEHLHLHYQENL 300  
 DB 449 HTNPLVSQILIRACKPYOLOIYFNERHSIRCPESGEHYEVLTLHFLQEYL 497

Search completed: October 15, 2003, 17:10:44  
 Job time : 19.9043 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 (Search time 6.91621 Seconds  
(without alignments)  
1896.467 Million cell updates/sec

Title: US-10-070-464-3  
Perfect score: 1660  
Sequence: 1 FEGTKDSPLHHLVYVSYVN.....HLHYLOENLGSRIALAKVI 310

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1680	100.0	310	4 US-09-794-236-4	Sequence 4, Appl1
2	337.5	20.1	766	4 US-10-002-593-6	Sequence 6, Appl1
3	333.5	19.9	755	5 PCT-US93-07923-3	Sequence 3, Appl1
4	333.5	19.9	759	5 PCT-US93-07923-2	Sequence 2, Appl1
5	333.5	19.9	766	1 US-08-230-491A-3	Sequence 3, Appl1
6	333.5	19.9	766	1 US-08-619-280A-3	Sequence 3, Appl1
7	333.5	19.9	766	2 US-08-940-391-3	Sequence 3, Appl1
8	333.5	19.9	766	4 US-09-794-236-1	Sequence 1, Appl1
9	327	19.5	771	4 US-09-462-284-2	Sequence 2, Appl1
10	302	18.0	760	1 US-08-230-491A-2	Sequence 2, Appl1
11	302	18.0	760	1 US-08-619-280A-2	Sequence 2, Appl1
12	302	18.0	760	2 US-08-940-391-2	Sequence 2, Appl1
13	207	12.3	632	3 US-09-016-080-1	Sequence 1, Appl1
14	152.5	9.1	657	4 US-09-355-166-1	Sequence 1, Appl1
15	147	8.8	721	4 US-09-390-234-20	Sequence 20, Appl1
16	143.5	8.5	622	2 US-08-664-646A-2	Sequence 2, Appl1
17	143.5	8.5	622	2 US-09-066-285-2	Sequence 2, Appl1
18	143.5	8.5	622	2 US-09-261-006-2	Sequence 2, Appl1
19	143.5	8.5	622	3 US-08-951-088-2	Sequence 2, Appl1
20	143.5	8.5	622	4 US-09-609-566-2	Sequence 2, Appl1
21	143.5	8.5	622	4 US-09-609-570-2	Sequence 2, Appl1
22	143.5	8.5	622	4 US-09-427-372-2	Sequence 2, Appl1
23	143.5	8.5	622	4 US-09-693-554-2	Sequence 2, Appl1
24	138	8.2	614	4 US-09-252-991A-20060	Sequence 2, Appl1
25	110	6.5	275	4 US-09-198-452A-181	Sequence 181, App
26	95.5	5.7	344	4 US-09-724-623-79	Sequence 79, Appl
27	95	5.7	314	4 US-09-355-166-17	Sequence 17, Appl

28	93.5	5.6	256	4 US-09-355-166-2	Sequence 2, Appl1
29	91.5	5.4	570	3 US-09-068-960-13	Sequence 13, Appl
30	90.5	5.4	462	4 US-09-134-001C-4300	Sequence 4300, Ap
31	88.5	5.3	255	4 US-09-355-166-3	Sequence 3, Appl1
32	87	5.2	318	4 US-09-601-027-1	Sequence 1, Appl1
33	87	5.2	724	4 US-09-252-991A-30884	Sequence 30884, A
34	85.5	5.1	317	1 US-07-688-299-1	Sequence 1, Appl1
35	85.5	5.1	317	1 US-07-980-517A-1	Sequence 1, Appl1
36	85.5	5.1	318	1 US-07-688-299-3	Sequence 3, Appl1
37	85.5	5.1	318	1 US-07-688-299-13	Sequence 13, Appl
38	85	5.1	529	4 US-09-252-991A-24711	Sequence 24711, A
39	84	5.0	493	1 US-07-615-448A-7	Sequence 7, Appl1
40	84	5.0	493	1 US-08-196-361-7	Sequence 7, Appl1
41	84	5.0	493	2 US-08-446-934-7	Sequence 7, Appl1
42	84	5.0	493	2 US-08-448-128-7	Sequence 7, Appl1
43	84	5.0	493	3 US-08-948-703-7	Sequence 7, Appl1
44	83.5	5.0	686	3 US-09-368-169-8	Sequence 8, Appl1
45	82	4.9	414	4 US-09-107-532A-5022	Sequence 5022, Ap

## ALIGNMENTS

```
RESULT 1
US-09-794-236-4
; Sequence 4, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794.236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-4
```

Query Match 100.0%; Score 1680; DB 4; Length 310;  
Best Local Similarity 100.0%; Pred. No. 3.2e-174;  
Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	FEGTKDSPLHHLVYVSYVNGEYTRLTDRGSHSCCISOHCDFFISKYSNOKNPHCVSL	60
DB	1	FEGTKDSPLHHLVYVSYVNGEYTRLTDRGSHSCCISOHCDFFISKYSNOKNPHCVSL	60
QY	61	YKLSPEDDPCTKTEFPAATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOG	120
DB	61	YKLSPEDDPCTKTEFPAATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOG	120
QY	121	KYPTVPLTYGGPGQIETIDQVBSGLQYLASRYDFIDDRYGIHGSYGCTISLMAIMQR	180
DB	121	KYPTVPLTYGGPGQIETIDQVBSGLQYLASRYDFIDDRYGIHGSYGCTISLMAIMQR	180
QY	181	SDIPVATAGAPVTLMIYFDYGTERTYVGHDDNQGYYLDSVMAQAKFSEPRRLLL	240
DB	181	SDIPVATAGAPVTLMIYFDYGTERTYVGHDDNQGYYLDSVMAQAKFSEPRRLLL	240
QY	241	HGFLDENVHFATSIILSLFVRACKPYDQIYPOERHSIRVPESGEHETLHLHYLOENL	300
DB	241	HGFLDENVHFATSIILSLFVRACKPYDQIYPOERHSIRVPESGEHETLHLHYLOENL	300
QY	301	GSRIALAKVI 310	
DB	301	GSRIALAKVI 310	

RESULT 2

```
US-10-002-593-6
; Sequence 6, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTIN
; TITLE OF INVENTION: INHIBITOR AND VASOPEPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
; FILE REFERENCE: Acty Docket No. 6586198 1242/48/2
; CURRENT APPLICATION NUMBER: US/10/002,593
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/244,524
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO: 6
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-593-6
```

Query Match 20.1%; Score 337.5; DB 4; Length 766;

Best Local Similarity 28.0%; Pred. No. 1.3e-27;

Matches 99; Conservative 53; Mismatches 127; Indels 75; Gaps 13;

```
QY 5 KDSPLEHLVYVSYNPGSEVTRLTDRGYSHSCCIS-QHCDFFISKYSNQKNPH---C--- 57
DB 423 KMPGGRNLKYIQLIDYTKVTCI-----SCELNPERCQYVSFSFEAKYQLRCSGP 475
QY 58 -VSLYKLSPEDDPCTCKTEFWATILDSAGPLPDYTPPE---IFSFEFTGFTLYGLMYK 113
DB 476 GLPLTLHSSVNDKGLRVLED-NSALDKM--LQNVQMPKSLDFTIILNETKF-WYQMIIP 531
QY 114 PHDLOPKKKYPTVLFYIGP----- 133
DB 532 PH-FPKSKKYPILLDVAAPGSCQKADTVFRILMWTYLASTENIIVASFGDGGSGYQDKI 590
QY 134 -----QGOIIEIDQVEGLQYLASRYDFILDRVGIGHMSYGGYLSMALMQRSDFPV 186
DB 591 MHAIRRLGTFEVEDQIEAARQF-SKMGFVNDKRIALIMGMSYGGYVTSVLGSGGVFC 649
QY 187 A1AGA PVTLMIFYDGYTERYMG--HPDQNEQGYLLGSVAMQAEKPPSPNRLLLHGL 244
DB 650 GIAVAPVSRWEYDVSYTERYMGLPTEPDNLDRHNSVTMSRAENF--KQVEYLLIHGTA 707
QY 245 DENVHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 708 DDNVHFOQSAQISKALVDVGVDFOAMWYTTDEDHGIASSTAHOIYTHMSHFIKQ 761
```

RESULT 3  
PCT-US93-07923-3

; Sequence 3, Application PC/TUS9307923

; GENERAL INFORMATION:

; APPLICANT: Morimoto, Chikao

; APPLICANT: Schlossman, Stuart F.

; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 MB

; OPERATING SYSTEM: IBM PS/2 Model 502 or 55SX

; SOFTWARE: WordPerfect (Version 5.0)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/07923

```
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Frazer, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3
```

Query Match 19.9%; Score 333.5; DB 5; Length 755;

Best Local Similarity 27.7%; Pred. No. 3.5e-27;

Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;

```
QY 5 KDSPLEHLVYVSYNPGSEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKNPH--- 56
DB 412 KMPGGRNLKYI-----QLSD-YTKVTCISCELNPERCQYVSFSFEAKYQLR 460
QY 57 C---VSLYKLSPEDDPCTCKTEFWATILDSAGPLPDYTPPE---IFSFEFTGFTLYG 109
DB 461 CSGPLPLTYLHSSVNDKGLRVLED-NSALDKM--LQNVQMPKSLDFTIILNETKF-WYQ 516
QY 110 MLKXPHDLOPKKKYPTVLFYIGP----- 133
DB 517 MLPPH-FDKSKKYPILLDVAAPGSCQKADTVFRILMWTYLASTENIIVASFGDGGSGYQ 575
QY 134 -----QGOIIEIDQVEGLQYLASRYDFILDRVGIGHMSYGGYLSMALMQRSDFPV 182
DB 576 GDKIMHAIRRLGTFEVEDQIEAARQF-SKMGFVNDKRIALIMGMSYGGYVTSVLGSGSG 634
QY 188 IFRVAJAGAPVTLMIFYDGYTERYMG--HPDQNEQGYLLGSVAMQAEKPPSPNRLLL 240
DB 635 VFKCGIAVAPVSRWEYDVSYTERYMGLPTEPDNLDRHNSVTMSRAENF--KQVEYLLI 692
QY 241 HGFUDENVHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 693 HGTADDNVHFOQSAQISKALVDVGVDFOAMWYTTDEDHGIASSTAHOIYTHMSHFIKQ 750
```

RESULT 4  
PCT-US93-07923-2

; Sequence 2, Application PC/TUS9307923

; GENERAL INFORMATION:

; APPLICANT: Morimoto, Chikao

; APPLICANT: Schlossman, Stuart F.

; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 MB

; OPERATING SYSTEM: IBM PS/2 Model 502 or 55SX

; SOFTWARE: WordPerfect (Version 5.0)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07923  
FILING DATE: 19930819  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/934,162  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: 07/832,211  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Frazer, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 00530/055002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 759  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US93-07923-2

Query Match 19.9%; Score 333.5; DB 5; Length 759;  
Best Local Similarity 27.7%; Pred. No. 3.5e-27;  
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;  
QY 5 KDSPLHLLVYVSYNPEBVTRLTRGYSHSCCIS-----QHCDFFISKYSNQKNPH--- 56  
DB 416 KMPGGRMLYKRI-----QLSD--YTKVTCLSCELNPERCQYVSFSKAYYQUR 464  
QY 57 C-----VSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPPE---ISFESTTGFTLYG 109  
DB 465 CSGPGLPLYTHSSVNDKGLRVLED-NSALDKM--LQNVQWPSKLDFTILNETKF-WYQ 520  
QY 110 MLYKPHDLQPKKPYTVLFYIGP----- 133  
DB 521 MLRPH-PDKSKKPYLLDDVYAGPCSQKADTVFRMLMATYLASTENIIVASFDGSGSYQ 579  
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLVRVGHGWSYGGYLSIMALMORS 182  
DB 580 GDKIMHAINRRLGFEVEDEJEAARQF-SKMGFVDNKRIAIWMSYGGYVSMVLGSGSG 638  
QY 183 IFRVALAGAPVTLWTFYDGTGTERYMG--HPDQNEQGYLLSVMAQAKPSEPRRLLL 240  
DB 639 VFKGCIAPVSRMEYDVSITERYMGLPTEPDNDHYRNSITWSRAENF--KQVEYLLI 696  
QY 241 HGFLENVHFAHTSILSLFLVRACKPYDLQIYPOERHSIRVPESEGHYELHLHYLOE 298  
DB 697 HGTADDNVHFGQSQNISKALVDVGVDFOAMMYTDEDHGIASSTAHQHITYHMSHFIRK 754

RESULT 5  
US-08-230-491A-3  
Sequence 3, Application US/08230491A  
Patent No. 5587299  
GENERAL INFORMATION:  
APPLICANT: Rectig, Wolfgang J.; Scanlan, Matthew J.;  
TITLE OF INVENTION: Gatin-Chessa, Pilar; Old, Lloyd J.  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR  
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USGS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: FELFE & LYNCH  
STREET: 805 THIRD AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT - ASC II  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/230,491A  
FILING DATE: 20-APRIL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5587299man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 330  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 766 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-230-491A-3

Query Match 19.9%; Score 333.5; DB 1; Length 766;  
Best Local Similarity 27.7%; Pred. No. 3.5e-27;  
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;  
QY 5 KDSPLHLLVYVSYNPEBVTRLTRGYSHSCCIS-----QHCDFFISKYSNQKNPH--- 56  
DB 423 KMPGGRMLYKRI-----QLSD--YTKVTCLSCELNPERCQYVSFSKAYYQUR 471  
QY 57 C-----VSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPPE---ISFESTTGFTLYG 109  
DB 472 CSGPGLPLYTHSSVNDKGLRVLED-NSALDKM--LQNVQWPSKLDFTILNETKF-WYQ 527  
QY 110 MLYKPHDLQPKKPYTVLFYIGP----- 133  
DB 528 MLRPH-PDKSKKPYLLDDVYAGPCSQKADTVFRMLMATYLASTENIIVASFDGSGSYQ 586  
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLVRVGHGWSYGGYLSIMALMORS 182  
DB 587 GDKIMHAINRRLGFEVEDEJEAARQF-SKMGFVDNKRIAIWMSYGGYVSMVLGSGSG 645  
QY 183 IFRVALAGAPVTLWTFYDGTGTERYMG--HPDQNEQGYLLSVMAQAKPSEPRRLLL 240  
DB 646 VFKGCIAPVSRMEYDVSITERYMGLPTEPDNDHYRNSITWSRAENF--KQVEYLLI 703  
QY 241 HGFLENVHFAHTSILSLFLVRACKPYDLQIYPOERHSIRVPESEGHYELHLHYLOE 298  
DB 704 HGTADDNVHFGQSQNISKALVDVGVDFOAMMYTDEDHGIASSTAHQHITYHMSHFIRK 761

RESULT 6  
US-08-619-280A-3  
Sequence 3, Application US/06619280A  
Patent No. 5767242  
GENERAL INFORMATION:  
APPLICANT: Zimmermann, Rainer; Park, John E.;  
TITLE OF INVENTION: Rectig, Wolfgang; Old, Lloyd J.  
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-619-280A-3
```

```

Query Match          19.9%: Score 333.5; DB 1; Length 766;
Best Local Similarity 27.7%: Pred. No. 3.5e-27;
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;
```

```

QY 5 KDSPLEHLYVSYVNPGEVTRLTDRGYSHSCIS-----QHCDFFISKYSNOKNPH--- 56
DB 423 KMPGGRNLYKI-----QLSD--YTKVTCLSCENLPERCOYVSFSSEANYQLR 471
QY 57 C-----VSLYKLSPEDDPTCKTEFMATILDSAGLPDYTPPE---IFSFEETGFTLYG 109
DB 472 CSGPLPLTYLTHSSVNDKGLRVLED--NSALDKM--LQNVQPSKLDLFIINETRF-WYQ 527
QY 110 MLKRPDHPGKRPYTVLFYIGP----- 133
DB 528 MLPPH-FDKSKKPYLLLDVYAGPCSQKADTVFRLNMTATYLASTENIIVASFQSGSGQ 586
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLRVGIHGSYGGYLSMALMQRSD 182
DB 587 GDKIMHAINRRLGTREVEVDQIEAARQF-SKMGFVNDKRIAIWMSYGGYVTSMLVSGSG 645
QY 183 IFRVAIAGAPVTLWIFDYGTGYTERYMG--HPDQNEGGYLLGSVAMQAEKFPSEPNRLLL 240
DB 646 VFKCGIAVAPSRMEYDSVTERYMGLPTEPDNLDHYNSTVMSRAENF--KQVEYLLI 703
QY 241 HGFLEDNVPHATSTILSLFVLRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 704 HGTADNVHFOOSAQISKALVDVGVDFOAMWYTTDEDHGIASSTAHOIYTHMSHFIRKQ 761
```

## RESULT 7

```

US-08-940-391-3
Sequence 3, Application US/08940391
Patent No. 5965373
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/940,391
FILING DATE: 01-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/619,280
FILING DATE: 18-MARCH-1996
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5965373man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-940-391-3
```

```

Query Match          19.9%: Score 333.5; DB 2; Length 766;
Best Local Similarity 27.7%: Pred. No. 3.5e-27;
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;
```

```

QY 5 KDSPLEHLYVSYVNPGEVTRLTDRGYSHSCIS-----QHCDFFISKYSNOKNPH--- 56
DB 423 KMPGGRNLYKI-----QLSD--YTKVTCLSCENLPERCOYVSFSSEANYQLR 471
QY 57 C-----VSLYKLSPEDDPTCKTEFMATILDSAGLPDYTPPE---IFSFEETGFTLYG 109
DB 472 CSGPLPLTYLTHSSVNDKGLRVLED--NSALDKM--LQNVQPSKLDLFIINETRF-WYQ 527
QY 110 MLKRPDHPGKRPYTVLFYIGP----- 133
DB 528 MLPPH-FDKSKKPYLLLDVYAGPCSQKADTVFRLNMTATYLASTENIIVASFQSGSGQ 586
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLRVGIHGSYGGYLSMALMQRSD 182
DB 587 GDKIMHAINRRLGTREVEVDQIEAARQF-SKMGFVNDKRIAIWMSYGGYVTSMLVSGSG 645
QY 183 IFRVAIAGAPVTLWIFDYGTGYTERYMG--HPDQNEGGYLLGSVAMQAEKFPSEPNRLLL 240
DB 646 VFKCGIAVAPSRMEYDSVTERYMGLPTEPDNLDHYNSTVMSRAENF--KQVEYLLI 703
QY 241 HGFLEDNVPHATSTILSLFVLRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 704 HGTADNVHFOOSAQISKALVDVGVDFOAMWYTTDEDHGIASSTAHOIYTHMSHFIRKQ 761
```

## RESULT 8

```

US-09-794-236-1
Sequence 1, Application US/09794236
Patent No. 6337069
GENERAL INFORMATION:
APPLICANT: Grouzmann, Eric
APPLICANT: Lacroix, Jean-Silvain
APPLICANT: Monod, Michel
TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
FILE REFERENCE: 81985/276823
CURRENT APPLICATION NUMBER: US/09/794,236
CURRENT FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 766
TYPE: PRT
ORGANISM: Homo sapiens
US-09-794-236-1
```

```

Query Match          19.9%: Score 333.5; DB 4; Length 766;
Best Local Similarity 27.7%: Pred. No. 3.5e-27;
```

Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;

QY 5 KOSPLEHHLYVSVNPGEVTRLTDGYSHSCIS-----QHCDFFISKYSNOKNPH---56  
 Db 423 KMGPGRNLYKTI-----QLSD--YTKVTLCSCLNPERCOYISVSSKAKYQLR 471  
 QY 57 C---VSLYKSSPEDDPTCKTEFWATILDSAGPLPDYTPPE---IFSFEFTGFTLYG 109  
 Db 472 CCGPGILYTLHSSVNDKGRVLED-NSALDKM--LQNVMPKSKDLFIILNETKF-WYQ 527  
 QY 110 MLYKPHDLOPKKYPTVLFIYGGP-----133  
 Db 528 MLIPRH-PDKSKKYPLLDVYAGPCQKADTVFLNWTATYLASTENIIVASFDGSGGY 586  
 QY 134 -----OQIEIDDOVEGLQYLASRYDFIDLRVGIHGMVSGYLSLMAIMQRSD 182  
 Db 587 GDKIMAINIRRLCTFEEDQIEAROP-SMGGFVNDKRIAIKWMVSGYVTSVVLGSGG 645  
 QY 183 IFRVALAGAVTLIMFYDTGYTERYMG--HPDNEOGYVLGSVAMQAEKPESEPNLLLL 240  
 Db 646 VFRGCIAPVAVSVSMWEYDYSYTERYMGLPTEPNLDHRYNSTWMSAENF--KQVEYLLI 703  
 QY 241 HGLDENVHFATSIILSFLVRAGKPYDLOIYPOERHSIVPESGEHYELHLHYLOE 298  
 Db 704 HGTADNVHFOQSAQISKALVDVGVDFOAMMYTDEDHGIASTAHOHIYTHMSHFTKQ 761

## RESULT 9

US-09-462-284-2  
 / Sequence 2, Application US/09462284  
 / Patent No. 6309868  
 / GENERAL INFORMATION:  
 / APPLICANT: Nestec S.A.  
 / APPLICANT: Monod, Michel  
 / APPLICANT: Douglas, Agnes  
 / APPLICANT: Afolter, Michael  
 / APPLICANT: Van Den Broek, Peter  
 / TITLE OF INVENTION: CLONING OF THE  
 / TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM  
 / TITLE OF INVENTION: ASPERGILLUS ORYZAE  
 / FILE REFERENCE: 8265-298  
 / CURRENT APPLICATION NUMBER: US/09/462,284  
 / CURRENT FILING DATE: 2000-01-03  
 / NUMBER OF SEQ ID NOS: 9  
 / SOFTWARE: FastSeq for Windows Version 4.0  
 / SEQ ID NO 2  
 / LENGTH: 771  
 / TYPE: PRT  
 / ORGANISM: Fungus  
 / US-09-462-284-2

Query Match 19.5%; Score 327; DB 4; Length 771;

Best Local Similarity 29.9%; Pred. No. 1.8e-26; Mismatches 126; Indels 76; Gaps 11;

Matches 103; Conservative 40; Mismatches 126; Indels 76; Gaps 11;

QY 4 TKDSPLNHLHYVSVNPGEVTRLTDGYS--HSCISQHCDFISKYSNOKNPHCVSLY 61  
 Db 410 TQHSRERLHLSY-STRATVPLVDYAAVWSASFANSGYILLTGPDVPIY-QELX 467  
 QY 62 KLSSEDDPTCKTEFWATILDSAGPL--PDYTPPEIFSFE--STTGFTLYGMLYKPHD 116  
 Db 468 TINS-----TKPL-RTITDAKYLEOIKDYALFNITYFELPLPSGFTLNMQLPFG 518  
 QY 117 LQPGKYPTVLFIYGGPQG-----136  
 Db 519 FSPDKRPIPLFTPYGPGAGOEYTKRMOALNFKAYVASDELEVYTWVDNRGTGFGRRK 578  
 QY 137 -----IBIDDOVEGLQYLASRYDFIDLRVGIHGMVSGYLSLMAIMQRSDIFRY 186  
 Db 579 RSAVTRQLGLEAEDQIYAQQ-AANIPIWDADHIGIWMVSGGYILSVLEDSAGFTLL 637  
 QY 187 ALAGAPVTLMIFDYDGYTERYMGHPQONEGYVLGSVAMQAEKPESEPNRLLHLHGFLDE 246

Db 638 GVITAVSDMRFYDSMYTERYMKTLSTNEGYETSAV-RKTDGFKNVEGGLIQHGTGD 636  
 QY 247 NWFHHTSILSLFLVRAGKPYDQIYPOERHSIVPESGEHYELH 291  
 Db 697 NWFHNSAALVDLMDGD-----VSPERKLSQWFTSDHGISYH 735

## RESULT 10

US-08-230-491A-2  
 / Sequence 2, Application US/08230491A  
 / Patent No. 5587299  
 / GENERAL INFORMATION:  
 / APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;  
 / APPLICANT: Gartin-Chesa, Pilar; Old, Lloyd J.  
 / TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR  
 / TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES  
 / TITLE OF INVENTION: THEREOF  
 / NUMBER OF SEQUENCES: 10  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: FELPE & LYNCH  
 / STREET: 805 THIRD AVENUE  
 / CITY: NEW YORK  
 / STATE: NEW YORK  
 / COUNTRY: USA  
 / ZIP: 10022  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE  
 / COMPUTER: IBM PS/2  
 / OPERATING SYSTEM: PC-DOS  
 / SOFTWARE: WORDPERFECT - ASC II  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/230,491A  
 / FILING DATE: 20-APRIL-1994  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Hanson, No. 5587299man D.  
 / REGISTRATION NUMBER: 30,946  
 / REFERENCE/DOCKET NUMBER: LUD 330  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (212) 688-9200  
 / TELEFAX: (212) 838-3884  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 760 amino acids  
 / TYPE: amino acid  
 / TOPOLOGY: linear  
 / US-08-230-491A-2

Query Match 18.0%; Score 302; DB 1; Length 760;

Best Local Similarity 25.5%; Pred. No. 9.3e-24; Mismatches 139; Indels 78; Gaps 11;

Matches 93; Conservative 55; Mismatches 139; Indels 78; Gaps 11;

QY 1 FEGTDSPLNHLHYVSVNPGEVTRLTDGYSH--CCISQH-----CDFISKYSNOKN 54  
 Db 401 FRVQDSLFYSSNEFEYGRNRIYRISIGYSPSKCVTCHLRKRCQYVTASPSDYAK 460  
 QY 55 PRCVSLY-----KLSSEDDPTCKTEFWATILDSAGPLD-----YTPPEIFSFEFTG 105  
 Db 461 YVALVCYGGIPRISTLHDERDQOE---IKILEENKELNALKNQLPKEBIKKLEVDEI 516  
 QY 106 TLYGMLYKPHDLOPKKYPTVLFIYGGP-----133  
 Db 517 TLMYKMLRPPQFDRSKKYPLLIQVGGPSCQSVSVFAVMWISYLSKSGMVIALVDGNG 576  
 QY 134 -----OQIEIDDOVEGLQYLASRYDFIDLRVGIHGMVSGYLSLMAIM 178  
 Db 577 TAFQGDKLLYAVVRKLGVEVEEQITAVRKPI-EMGFIDEKRIAIKWMVSGYVSSIALA 635  
 QY 179 QRSDFIRVALIAGAPVTLMIFDYDGYTERYMGHP--DQNEOGYVLGSVAMQAEKPESEPNR 236  
 Db 636 SGTGIFKCGIAPVAVSVSMWEYDYSYTERYMGLPYKDDNLEHNKSTVMARAEYFRVVD-- 693  
 QY 237 LLLHGLDENVHFATSIILSFLVRAGKPYDLOIYPOERHSIVPESG---EHYELHL 293



```

Db      577 TAFQCDKLLVAVYRKLGVEVEQITAVAKFI-EMGFIDEKRIALIMGMEYGGVSLALA 635
Qy      179 QRSDFRVAIAGAPVTLMIFYDTGTERMGHP--DNEQGYLLGSVAMQAKPSEPNR 236
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      636 SGTGPFKCGIAVAPVSSWEYVASYTERFMGLPTDNDLNEHKNSIVMAAEYFRVVD-- 693
Qy      237 LLLHGFDENVHFAHTSILLSFLVRAGKPYDLOIYPOERHSIRVDESG--EHEYELHL 293
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      694 YLLHGTADNVHFONSQAIKALVNAQVDFQAMVYSDQNHGL---SGLSTNHLTYTMT 749
Qy      294 HYLQE 298
        : : : :
Db      750 HFLKQ 754

```

## RESULT 13

```

US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshitsugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; EARLIER FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

```

```

Query Match      12.3%; Score 207; DB 3; Length 632;
Best Local Similarity 25.4%; Pred. No. 1,5e-13;
Matches 61; Conservative 38; Mismatches 79; Indels 62; Gaps 7;

Qy      94 PEIPEFESTGTFTLGMLYKPHDLQPGKKYPTVLFYGGPO----- 134
        ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db      375 PEHFFVKASDGEVIDAMWKPVNFRKGYPALEIHGGPKYAGYAFMEHFVLTSGKF 434
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      135 -----GOIEIDQVEGLQYLASRYDFIDLRVGIHGMSTYGG 170
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      435 VVIFSNPRGSDGYGEEFADIRGHYGERDYQDLMEEVVDALRRFDFIDGERLDVGTGSGYG 494
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      171 YLSMALMQRSDIFRVAIAGAPVTLMIFY---DTGYTERYMGHPDQ-----NEOGY 219
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      495 FMTNW-IVGHTNRFKAAVTORSISNWSIFGTTDICY--YPA-PDQIKDPWMSNLEGW 549
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      220 LGSVAMQAEKPPSEPNRLLLHGFDENVHFAHTSILLSFLVRAGKPYDLOIYPOERHSI 279
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      550 EKSPKLYA--PNVETPLTIHSTEDYRCWLPDALQLFLSKYLGKRVALLAFPENHDL 606
        : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 14

```

US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GCS11-PCT
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21

```

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1

```

```

Query Match      9.1%; Score 152.5; DB 4; Length 657;
Best Local Similarity 20.1%; Pred. No. 1.3e-07;
Matches 72; Conservative 48; Mismatches 121; Indels 117; Gaps 12;

```

```

Qy      4 TKDSPLEHLLVYVSYVNPGEVTRLDRG-----YSHSCCISQHC 42
        ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db      311 TKDS---QGFFYIG-----TDQSGTGIYISIEGLVYPIRLEKYNISFSLSPDE 357
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      43 DFFISKYSNQKPHCVSLKYLSSPEDDPTCKTEFWATILDSAGLPDYT--PPEIFSE 100
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      358 OHFIASVTKPDRPSSLYSIPLGOEKKOLTGANDKF-----VREHTISIPERIQYA 407
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      101 STTGFTLYGMLYKPHDLQPGKKYPTVLFYGGPO----- 134
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      408 TEDGVAMVGMWLMRPAQMESETTYPLILNHGGPHMYGHTYHEFOVYLAAGYAVVYINP 467
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      135 -----GOIEIDQVEGLQYLASRYDFIDLRVGIHGMSTYGYLSMA 176
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      468 RSHGYGEPVNAVAVGDKMDYDVQAVDEAIKRDPHIDPKRLGVGTGSGYGFMTNW- 526
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      177 LMQRSDIFRVAIAGAPVTLMIFY---DTGY--TERYMGHP-----DNEQGYIYG 221
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      527 IVGQTRRKAATORSISNWSIFHGVSDIGYFFTMOLEHMFEDTEKLMDSPLKY--- 583
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      222 SVAMQAEKPPSEPNRLLLHGFDENVHFAHTSILLSFLVRAGKPYDLOIYPOERHSI 279
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      584 --AANVE-----TPLLHGERDRCPLEQAEQLFIALKKMGKETKLVFPNASHNL 633
        : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 15

```

US-09-390-234-20
; Sequence 20, Application US/09390234
; Patent No. 6365390
; GENERAL INFORMATION:
; APPLICANT: Blum, David L.
; APPLICANT: Katsueva, Irina
; APPLICANT: Li, Xin-liang
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Phenolic Acid Esterases, Coding Sequences and Methods
; FILE REFERENCE: 67-98
; CURRENT APPLICATION NUMBER: US/09/390,234
; EARLIER FILING DATE: 1999-09-03
; EARLIER APPLICATION NUMBER: US 60/099,136
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 721
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-09-390-234-20

```

```

Query Match      8.8%; Score 147; DB 4; Length 721;
Best Local Similarity 20.6%; Pred. No. 6.1e-07;
Matches 58; Conservative 33; Mismatches 98; Indels 92; Gaps 8;

```

```

Qy      93 PEIPEFESTGTFTLGMLYKPHDLQPGKKYPTVLFYGGPO----- 136
        ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db      439 PSDISEFYQGNFTDIIHAMVIYPENFDKSKYPLFIHGGPQGMADGWSTRMPKAWA 498
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      137 -----IEIDQVEGLQYLASRYDFIDLRVGIHGMSTYGY 165
        : : : : : : : : : : : : : : : : : : : : : : : :
Db      489 DQGYVVAAPNPTGSGTGFQALTLTAQNMGGAAPDIDLKCEVYHENIDLVDTDGVAAAG 558
        : : : : : : : : : : : : : : : : : : : : : : : :
Qy      166 WSYGYL-----SLMALMQRSDIFRVAIAGAPVT--LWIFYDTGYTERYMGHP 211
        ||| : : : : : : : : : : : : : : : : : : : : : : : :

```



Db	559	ASVGGFMIMWIOSSPLGRKKALVSHDGT---	VADAKVSTELWFM-----	602
Qy	212	DONEQYVIGSVAMQAEK----	PSEPNR-----	602
Db	603	---QREFNGTWDARDNRRWDPSA	PERILOPATPMVLVHSDKDYRLPVAEGSLFNVL	658
Qy	261	VRAGKPYDLQIYPOERHSIRVPE	SGEYELHLHYLOENMG	301
Db	659	QERGVPSSRPLNFPDENHWVVPEN	SLVWHQALGWINKYSG	699

Search completed: October 15, 2003, 17:15:35  
Job time : 7.91621 secs



Db 181 SDIFRVAIAGAPVTLMIIFYDTGTYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLL 240  
 QY 241 HGFLENVFAHTSILSLFLVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENTL 300  
 Db 241 HGFLENVFAHTSILSLFLVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENTL 300  
 QY 301 GSRIALAKVI 310  
 Db 301 GSRIALAKVI 310

## RESULT 2

US-09-976-674-1  
 ; Sequence 1, Application US/09976674  
 ; Patent No. US20020115843A1  
 ; GENERAL INFORMATION:

APPLICANT: Qi, Steve  
 APPLICANT: Akimsanya, Karen  
 APPLICANT: Riviere, Pierie  
 APPLICANT: Julien, Jean-Louis  
 TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV  
 FILE REFERENCE: 70669  
 CURRENT APPLICATION NUMBER: US/09/976,674  
 CURRENT FILING DATE: 2001-10-12  
 PRIOR APPLICATION NUMBER: US 60/240,117  
 PRIOR FILING DATE: 2000-10-12  
 NUMBER OF SEQ ID NOS: 61  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 1  
 LENGTH: 882  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-976-674-1

Query Match 97.9%; Score 1645.5; DB 10; Length 882;  
 Best Local Similarity 86.4%; Pred. No. 4e-161; Mismatches 0; Indels 49; Gaps 1;  
 Matches 310; Conservative 0;

QY 1 FEGTDSPLEHLLYVSYVNPGEVTRLTDGYSHSCCISOHCDFFISKYSNOKNPHCVSL 60  
 Db 524 FEGTDSPLEHLLYVSYVNPGEVTRLTDGYSHSCCISOHCDFFISKYSNOKNPHCVSL 583  
 QY 61 YKLSSEDDPTCKTKEFMATILDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 120  
 Db 584 YKLSSEDDPTCKTKEFMATILDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 643  
 QY 121 KKPPTVLFYGGPO----- 134  
 Db 644 KKPPTVLFYGGPOVQVOLLNRRFKGVKFRPLNTLASLGIYVVVVIDNRGSGHGLKEGAFK 703  
 QY 135 ---GQIEIDDOVEGLOYLASRYDFIDLDRVGIGHMSYGGYLSMALMORSDFRVAIAGA 191  
 Db 704 YKMGQIEIDDOVEGLOYLASRYDFIDLDRVGIGHMSYGGYLSMALMORSDFRVAIAGA 763  
 QY 192 PVTLMIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 251  
 Db 764 PVTLMIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 823  
 QY 252 HTSILSLFLVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENTGSRIALAKVI 310  
 Db 824 HTSILSLFLVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENTGSRIALAKVI 882

## RESULT 3

US-10-054-776-2  
 ; Sequence 2, Application US/10054776  
 ; Publication No. US20030165818A1  
 ; GENERAL INFORMATION:

APPLICANT: Mark Robert Edbrooke  
 APPLICANT: Alan Peter Lewis  
 TITLE OF INVENTION: NOVEL PROTEIN  
 FILE REFERENCE: OG1042US  
 CURRENT APPLICATION NUMBER: US/10/054,776

; CURRENT FILING DATE: 2002-01-23  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 882  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-054-776-2

Query Match 97.9%; Score 1645.5; DB 12; Length 882;  
 Best Local Similarity 86.4%; Pred. No. 4e-161; Mismatches 0; Indels 49; Gaps 1;  
 Matches 310; Conservative 0;

QY 1 FEGTDSPLEHLLYVSYVNPGEVTRLTDGYSHSCCISOHCDFFISKYSNOKNPHCVSL 60  
 Db 524 FEGTDSPLEHLLYVSYVNPGEVTRLTDGYSHSCCISOHCDFFISKYSNOKNPHCVSL 583  
 QY 61 YKLSSEDDPTCKTKEFMATILDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 120  
 Db 584 YKLSSEDDPTCKTKEFMATILDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 643  
 QY 121 KKPPTVLFYGGPO----- 134  
 Db 644 KKPPTVLFYGGPOVQVOLLNRRFKGVKFRPLNTLASLGIYVVVVIDNRGSGHGLKEGAFK 703  
 QY 135 ---GQIEIDDOVEGLOYLASRYDFIDLDRVGIGHMSYGGYLSMALMORSDFRVAIAGA 191  
 Db 704 YKMGQIEIDDOVEGLOYLASRYDFIDLDRVGIGHMSYGGYLSMALMORSDFRVAIAGA 763  
 QY 192 PVTLMIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 251  
 Db 764 PVTLMIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 823  
 QY 252 HTSILSLFLVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENTGSRIALAKVI 310  
 Db 824 HTSILSLFLVRAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOENTGSRIALAKVI 882

## RESULT 4

US-10-170-789-38  
 ; Sequence 38, Application US/10170789  
 ; Publication No. US20030180930A1  
 ; GENERAL INFORMATION:

APPLICANT: Rachel E. Meyers  
 APPLICANT: Oladett, Peter J.  
 APPLICANT: Kapeller-Libermann, Rosana  
 APPLICANT: Curtis, Rory A. J.  
 APPLICANT: Williamson, Mark  
 APPLICANT: Welch, Nadine  
 TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
 FILE REFERENCE: 10448-191001  
 CURRENT APPLICATION NUMBER: US/10/170,789  
 CURRENT FILING DATE: 2002-06-13  
 PRIOR APPLICATION NUMBER: US 09/797,039  
 PRIOR FILING DATE: 2001-02-28  
 PRIOR APPLICATION NUMBER: PCT/US01/06525  
 PRIOR FILING DATE: 2001-02-28  
 PRIOR APPLICATION NUMBER: US 60/186,061  
 PRIOR FILING DATE: 2000-02-29  
 PRIOR APPLICATION NUMBER: US 09/882,166  
 PRIOR FILING DATE: 2001-06-15  
 PRIOR APPLICATION NUMBER: PCT/US01/19269  
 PRIOR FILING DATE: 2001-06-15  
 PRIOR APPLICATION NUMBER: US 60/212,078  
 PRIOR FILING DATE: 2000-06-15  
 PRIOR APPLICATION NUMBER: US 09/934,406  
 PRIOR FILING DATE: 2001-08-21  
 PRIOR APPLICATION NUMBER: PCT/US01/26052  
 PRIOR FILING DATE: 2001-08-21  
 PRIOR APPLICATION NUMBER: US 60/226,740  
 PRIOR FILING DATE: 2000-08-21  
 PRIOR APPLICATION NUMBER: US 09/861,801

```
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: PCT/US01/16549
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: US 60/205,508
/ PRIOR FILING DATE: 2000-05-19
/ PRIOR APPLICATION NUMBER: US 09/801,267
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07138
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,454
/ PRIOR FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: US 09/829,671
/ PRIOR FILING DATE: 2001-04-10
/ PRIOR APPLICATION NUMBER: PCT/US01/40483
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: US 60/197,508
/ PRIOR FILING DATE: 2000-04-18
/ PRIOR APPLICATION NUMBER: US 09/961,721
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: PCT/US01/29904
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: US 60/235,023
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US 10/045,367
/ PRIOR FILING DATE: 2001-11-07
/ PRIOR APPLICATION NUMBER: US 60/246,561
/ PRIOR FILING DATE: 2000-11-07
/ PRIOR APPLICATION NUMBER: US 09/801,275
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07074
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,420
/ PRIOR FILING DATE: 2000-03-07
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 38
/ LENGTH: 882
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-170-789-38
```

```
Query Match      97.9%; Score 1645.5; DB 12; Length 882;
Best Local Similarity 86.4%; Pred. No. 4e-161;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;
```

```
QY      1 FEGTDSPLEHLLVYVSYVNPGEVTRLDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 60
DB      524 FEGTDSPLEHLLVYVSYVNPGEVTRLDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 583
QY      61 YLSSPEDDPCTCKTEFNATLIDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOG 120
DB      584 YLSSPEDDPCTCKTEFNATLIDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOG 643
QY      121 KKYPTLVFIYGGPQ----- 134
DB      644 KKYPTLVFIYGGPQVQLVNNRFKGVKYPRLNTLASLGVAVVVIDNRGSCRGKKEGAFK 703
QY      135 ---GQIEIDDOVEGQIYASRYDFIDLDRVGHGMSYGGYSLMALMORSDFRVAIAGA 191
DB      704 YKMGIEIDDOVEGQIYASRYDFIDLDRVGHGMSYGGYSLMALMORSDFRVAIAGA 763
QY      192 PVTLMIFPDGTGYTERYMGHPDNEQGYLYGSVAMQAEKPPSEPNLLLLHGFLENVHFA 251
DB      764 PVTLMIFPDGTGYTERYMGHPDNEQGYLYGSVAMQAEKPPSEPNLLLLHGFLENVHFA 823
QY      252 HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYQENLGSRIALAKYI 310
DB      824 HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYQENLGSRIALAKYI 882
```

```
RESULT 5
US-09-976-674-3
/ Sequence 3, Application US/09976674
```

```
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Qi, Steve
/ APPLICANT: Akinsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Junten, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976,674
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 863
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-976-674-3
```

```
Query Match      65.1%; Score 1093.5; DB 10; Length 863;
Best Local Similarity 57.3%; Pred. No. 5.3e-104;
Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;
```

```
QY      1 FEGTDSPLEHLLVYVSYVNPGEVTRLDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 60
DB      515 FEGTDSPLEHLLVYVSYVNPGEVTRLDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 574
QY      61 YLSSPEDDPCTCKTEFNATLIDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOG 120
DB      575 YLSSPEDDPCTCKTEFNATLIDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOG 634
QY      121 KKYPTLVFIYGGPQ----- 134
DB      635 KKYPTLVFIYGGPQVQLVNNRFKGVKYPRLNTLASLGVAVVVIDNRGSCRGKKEGAFK 694
QY      135 ---GQIEIDDOVEGQIYASRYDFIDLDRVGHGMSYGGYSLMALMORSDFRVAIAGA 191
DB      695 NQMGVEIEDDOVEGQIYASRYDFIDLDRVGHGMSYGGYSLMALMORSDFRVAIAGA 754
QY      192 PVTLMIFPDGTGYTERYMGHPDNEQGYLYGSVAMQAEKPPSEPNLLLLHGFLENVHFA 251
DB      755 PVTLMIFPDGTGYTERYMGHPDNEQGYLYGSVAMQAEKPPSEPNLLLLHGFLENVHFA 814
QY      252 HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYQENL 300
DB      815 HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYQENL 863
```

```
RESULT 6
US-09-976-674-23
/ Sequence 23, Application US/09976674
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Qi, Steve
/ APPLICANT: Akinsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Junten, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976,674
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 23
/ LENGTH: 892
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-976-674-23
```

```
Query Match      65.1%; Score 1093.5; DB 10; Length 892;
```



PRIOR FILING DATE: 2000-10-12  
 NUMBER OF SEQ ID NOS: 61  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO: 35  
 LENGTH: 879  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-976-674-35

Query Match 64.7%; Score 1086.5; DB 10; Length 879;  
 Best Local Similarity 57.2%; Pred. No. 2.9e-103;  
 Matches 199; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 2 EGTGDSPLFHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 61  
 DB 532 KOTKOTPLEHHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 591  
 QY 62 KLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTGLYGMLYKPHDLQPGK 121  
 DB 592 KLSGDDDDPLHKQPRFMSAMEAASCPDPYVPEIFHFTFRSDVRLYGMVYKPHALQPGK 651  
 QY 122 KYPTVLFYIGGPQ----- 134  
 DB 652 KHPVLFYVGGPOVOLVNNSEFKIKYLRNLTLASLGAVVVIDRGSCQGRLEFGALKN 711  
 QY 135 --GQIEIDQVYEGLOYLASRYDFIDLDRVGIHMSYGYLSLMALMORSDFRVAIAGAP 192  
 DB 712 QMGQVEIEDQVYEGLOFVAEKYGFIDLSRVAIHGMSYGGFLSMGLIHKQVFKVIAIAGAP 771  
 QY 193 VTLWIFDYDTGYTERYMGHPDQNEOGYVYLGSAVMAQEKPESEPNRLLLHGLDENVHPAH 252  
 DB 772 VTVWMAVDTGYTERYMDVPENNQHYEAGSVALLHVEKLPNEPNRLLLHGLDENVHPHF 831  
 QY 253 TSILSLFVLRAGKRPYDLOIYPERHSIRPESGEHYELHLHYLOENL 300  
 DB 832 TNFLVSQILIRAGKPYOLOIYPERHSIRCPESGEHYEVLHLHFLQEYL 879

RESULT 10  
 US-10-237-271-3  
 Sequence 3, Application US/10237271  
 Publication No. US20030096328A1  
 GENERAL INFORMATION:  
 APPLICANT: THE BURHAM INSTITUTE  
 APPLICANT: SMITH, Jeffrey W.  
 APPLICANT: KRIDEL, Steven J.  
 APPLICANT: AXELROD, Pamiho T.  
 TITLE OF INVENTION: SERINE/THREONINE HYDROLASE PROTEINS AND SCREENING ASSAYS  
 FILE REFERENCE: BURN1100-1  
 CURRENT APPLICATION NUMBER: US/10/237,271  
 PRIOR FILING DATE: 2002-09-04  
 PRIOR APPLICATION NUMBER: US 60/317,842  
 PRIOR FILING DATE: 2001-09-06  
 NUMBER OF SEQ ID NOS: 19  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 3  
 LENGTH: 508  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-237-271-3

Query Match 61.5%; Score 1032.5; DB 15; Length 508;  
 Best Local Similarity 54.2%; Pred. No. 5.1e-98;  
 Matches 194; Conservative 43; Mismatches 54; Indels 67; Gaps 3;

QY 1 FEGTKDSPLEHHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 60  
 DB 160 FQGTGDPLEHHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 219  
 QY 61 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTGLYGMLYKPHDLQPG 120  
 DB 220 YKLSGDDDDPLHKQPRFMSAMEA-----KIFHFTRSDVRLYGMVYKPHALQPG 270

QY 121 KYPTVLFYIGGPQ----- 134  
 DB 271 KHPVLFYVGGPOVOLVNNSEFKIKYLRNLTLASLGAVVVIDRGSCQGRLEFGALKN 330  
 QY 135 --GQIEIDQVYEGLOYLASRYDFIDLDRVGIHMSYGYLSLMALMORSDFRVAIAGAP 191  
 DB 331 NQMGVEIEDQVYEGLOFVAEKYGFIDLSRVAIHGMSYGGFLSMGLIHKQVFKVIAIAGAP 390  
 QY 192 VTLWIFDYDTGYTERYMGHPDQNEOGYVYLGSAVMAQEKPESEPNRLLLHGLDENVHPAH 251  
 DB 391 VTVWMAVDTGYTERYMDVPENNQHYEAGSVALLHVEKLPNEPNRLLLHGLDENVHPHF 450  
 QY 252 TSILSLFVLRAGKRPYDLOIYPERHSIRPESGEHYELHLHYLOENL 300  
 DB 451 TNFLVSQILIRAGKPYOLOIYPERHSIRCPESGEHYEVLHLHFLQEYL 508

RESULT 11  
 US-09-976-674-19  
 Sequence 19, Application US/09976674  
 Patent No. US20020115843A1  
 GENERAL INFORMATION:  
 APPLICANT: Qi, Steve  
 APPLICANT: Akinsanya, Karen  
 APPLICANT: Riviere, Pierre  
 APPLICANT: Julien, Jean-Louis  
 TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V  
 FILE REFERENCE: 70669  
 CURRENT APPLICATION NUMBER: US/09/976,674  
 CURRENT FILING DATE: 2001-10-12  
 PRIOR APPLICATION NUMBER: US 60/240,117  
 PRIOR FILING DATE: 2000-10-12  
 NUMBER OF SEQ ID NOS: 61  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 19  
 LENGTH: 658  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-976-674-19

Query Match 44.0%; Score 739; DB 10; Length 658;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-67;  
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEGTKDSPLEHHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 60  
 DB 524 FEGTKDSPLEHHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 583  
 QY 61 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTGLYGMLYKPHDLQPG 120  
 DB 584 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTGLYGMLYKPHDLQPG 643  
 QY 121 KYPTVLFYIGG 132  
 DB 644 KYPTVLFYIGG 655

RESULT 12  
 US-09-976-674-11  
 Sequence 11, Application US/09976674  
 Patent No. US20020115843A1  
 GENERAL INFORMATION:  
 APPLICANT: Qi, Steve  
 APPLICANT: Akinsanya, Karen  
 APPLICANT: Riviere, Pierre  
 APPLICANT: Julien, Jean-Louis  
 TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V  
 FILE REFERENCE: 70669  
 CURRENT APPLICATION NUMBER: US/09/976,674  
 CURRENT FILING DATE: 2001-10-12  
 PRIOR APPLICATION NUMBER: US 60/240,117  
 PRIOR FILING DATE: 2000-10-12  
 NUMBER OF SEQ ID NOS: 61

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 661
; TYPE: PR1
; ORGANISM: Homo sapiens
; US-09-976-674-11

```

Query Match	44.0%;	Score 739;	DB 10;	Length 661;
Best Local Similarity	100.0%;	Pred. No. 1.8e-67;		
Matches 132;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	FECTKDSPLEHNLVYVSYVNPGEVTRLDRGYSHSCCISOHDFISXYSQKMPHCYSL	60
Db	524	FECTKDSPLEHNLVYVSYVNPGEVTRLDRGYSHSCCISOHDFISXYSQKMPHCYSL	583
Qy	61	YKLSSEDDPCTCKTFMAITLDSAGPLADYPRPIFSFESTGTGLYGMLYKHNDLQPG	120
Db	584	YKLSSEDDPCTCKTFMAITLDSAGPLADYPRPIFSFESTGTGLYGMLYKHNDLQPG	643
Qy	121	KKPTVLFIYGG	132
Db	644	KKPTVLFIYGG	655

```

RESULT 13
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;
; APPLICANT: O'i Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Julien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
;
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: patentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-976-674-7

```

QY	1	FEGTKDSPLEHNLVYVYNPGEVTRLTDROGYSHSCCISQHDPEFLSKYSKNGKHPACVSL	60
Db	524	FEGTKDSPLEHNLVYVYNPGEVTRLTDROGYSHSCCISQHDPEFLSKYSKNGKHPACVSL	583
QY	61	YKUSSPEDDPCKTKEFWATLLDSAGRLPDYTPPEIFSESESTGTLLYGMLYKPHDLOG	120
Db	584	YKUSSPEDDPCKTKEFWATLLDSAGRLPDYTPPEIFSESESTGTLLYGMLYKPHDLOG	643
QY	121	KKYPVLFIYGG	132
Db	644	KKYPVLFIYGG	655

RESULT 14  
US-09-976-674-29  
Sequence 29, Application US/09976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Akisanya, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Juntten, Jean-Louis

```

: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
:
: FILE REFERENCE: 70669
:
: CURRENT APPLICATION NUMBER: US/09/976,674
:
: CURRENT FILING DATE: 2001-10-12
:
: PRIOR APPLICATION NUMBER: US 60/240,117
:
: PRIOR FILING DATE: 2000-10-12
:
: NUMBER OF SEQ ID NOS: 61
:
: SOFTWARE: Patentin version 3.1
:
: SEQ ID NO 29
:
: LENGTH: 832
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: US-09-976-674-29

```

[illegible]

```

RESULT 15
US-09-976-674-31
/ Sequence 31, Application US/09976674
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Ol, Steve
/ APPLICANT: Akhsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Jundien, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976, 674
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 31
/ LENGTH: 832
/ TYPE: PR1
/ ORGANISM: Homo sapiens
/ US-09-976-674-31

```

[illegible]

QY	121	KKYPTVLFYGGPQ-----	134
Db	664	KHHPTVLFYGGPQVOVLNNSFKGIXYLRNLTLASLGYAVVWDGRGSCQRGILRFGALK	723
QY	135	---GQIEIDDQVEGLQYLASRYDIDLRVGIHGWMSYGYLSLMALMORSDFRVAIAGA	191
Db	724	NQMGQVEIEDQVEGLQFVAEKYGFIDLSRVVAIHGWMSYGFSLMGLIHKPQVFKQOPLAY	783
QY	192	PVTL	195
Db	784	PPRL	787

Search completed: October 15, 2003, 17:28:49  
Job time : 44.8027 secs





## RESULT 2

AB7516  
 dipeptidyl peptidase IV (Imported) - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: AB7516  
 R.Nietman, W.C.; Fejldjy, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heideberg, J.B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolotn, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: AB7249; MUID:21173698; PMID:11259647  
 A/Accession: AB7516  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-738 <STO>  
 A/Cross-references: GB:AE005673; NID:g13423647; PIDN:AAK24125.1; GSPDB:GNO0148  
 C/Genetics:  
 A/Gene: CC2154

Query Match 28.8%; Score 484.5; DB 2; Length 738;  
 Best Local Similarity 32.2%; Pred. No. 5e-34;  
 Matches 106; Conservative 53; Mismatches 109; Indels 61; Gaps 6;  
 QY 1 FEETKDSPLEHLLVYVSVNPEVTRLTDRGYSHSCISQHCDFIFISKYSNOKNPHCVSL 60  
 DB 394 FMSITPIERILYEVSYAKPKKALTSAGGMMMAKAVADNGAFACTYSDPTPSOTAL 453  
 QY 61 YKLSSP-----EDPTCKTEFWATILDSAGPLPDYTPPEIFSESTTGYMLYKPH 115  
 DB 454 YSADGKRWIRIENKLAEGHPW----PYANLPQ---PFGSLKADGETLHYELKPI 506  
 QY 116 DLQPGKKYPTVLFIYGGPQ----- 134  
 DB 507 GFDPACKYPAIVSVGGPHAQRMKNHSPSEERTYEAGYVIFKLDNRSGNSAKFMRA 566  
 QY 135 -----COIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYSLSMALMQRSDIFRVALA 189  
 DB 567 LDRKLGTIVEVEDLLGAKFLASQ-PYVDADKLQMGMSYSGFMALMLLTMENTPFRAGAA 625  
 QY 190 GAPTTLWIFYDTGTERYMGHPDQNEGYVLGSVAMQAEKFPSEPNRLLLHGFLENVH 249  
 DB 626 GAPTETSLYDTAYTERYMGKPDENKAGYVSDINNRDL--APGSLLLHGMADDNVI 683  
 QY 250 FAHTSILSLFLVRAKGYDQIYPOERHS 278  
 DB 684 FENSTRMLAALQKRAILFEMAMPGERHS 712

## RESULT 3

JC5142  
 X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia  
 C/Species: Xanthomonas maltophilia  
 C/Date: 31-Jan-1997 #sequence\_revision 27-Feb-1997 #text\_change 20-Jun-2000  
 C/Accession: JC5142  
 R.Kabashima, T.; Ito, K.; Yoshimoto, T.  
 J. Biochem. 120, 1111-1117, 1996  
 A/Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expression  
 A/Reference number: JC5142; MUID:97164011; PMID:9010758  
 A/Accession: JC5142  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-741 <KAB>  
 A/Cross-references: DDBJ:D83263; NID:g1753196; PIDN:BA11872.1; PID:g1753197  
 A/Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl residues from proteins.  
 C/Superfamily: dipeptidyl-peptidase IV  
 C/Keywords: dipeptidylpeptidase; hydrolase; membrane bound  
 F/4-18/Domain: transmembrane #status predicted <TM>  
 F/610/Active site: Ser #status predicted  
 F/685/Active site: Asp #status predicted  
 F/717/Active site: His #status predicted

Query Match 23.7%; Score 398.5; DB 2; Length 741;  
 Best Local Similarity 28.0%; Pred. No. 1.5e-26;  
 Matches 98; Conservative 57; Mismatches 120; Indels 75; Gaps 8;

QY 1 FEETKDSPLEHLLVYVSVNPEVTRLTDRGYSHSCISQHCDFIFISKYSNOKNPHCVSL 60  
 DB 396 FRAGIRASRQSLYAVP-LQGQOPQLSKAPKMSHSAFARNASVYVSDSNSTPQIQL 454  
 QY 61 YKLSSPDDPTCKTEFWATIL--DSAGLPDYT-----PEIISFESTTGYT-LYGM 110  
 DB 455 FRANG-----EKITLVENDLADKHPARYAREARQPFVETGLTAADGKTPLNYS 504  
 QY 111 LYKPEHLOGKKYPTVLFIYGGP----- 133  
 DB 505 VIKRAGFDPACKYPAIVSVYGGPASTVTYDMSWGRGDHLFNOYLAQGVVFLDNRGTP 564  
 QY 134 -----OQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYSLSMALMQR 180  
 DB 565 RRGDFGALYKQGTVEVADQLRGVAML-KQPPWDPARIGVQWNSNGYMTLMALAKA 623  
 QY 181 SDIFRAIAGAPYTLWIFYDTGTERYMGHPDQNEGYVLGSVAMQAEKFPSEPNRLLL 240  
 DB 624 SDSYACGAVAGAPYDGLYDHYTERYMDLPARNAGYREARVLTIEGLRSP--LLLI 680  
 QY 241 HGFLENVHFAHTSILSLFLVRAKGYDQIYPOERHSIRVPSGHEYL 290  
 DB 681 HGMAADNVLTSTSLMSALQKRGQPFELMTYFGAGHGSLGADALHRYV 730

## RESULT 4

S66261  
 X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum  
 C/Species: Flavobacterium meningosepticum  
 C/Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 20-Jun-2000  
 C/Accession: S66261  
 R.Kabashima, T.; Ito, K.; Yoshimoto, T.  
 Arch. Biochem. Biophys. 320, 123-128, 1995  
 A/Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from F.  
 A/Reference number: S66261; MUID:95314307; PMID:7793970  
 A/Accession: S66261  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-711 <KAB>  
 A/Cross-references: EMBL:D42121; NID:g577283; PIDN:BAA07702.1; PID:g577284  
 C/Superfamily: dipeptidyl-peptidase IV  
 C/Keywords: dipeptidylpeptide hydrolase

Query Match 22.6%; Score 379; DB 2; Length 711;  
 Best Local Similarity 26.6%; Pred. No. 7.2e-25;  
 Matches 90; Conservative 58; Mismatches 118; Indels 72; Gaps 5;  
 QY 19 VNPGEVTRLTDRGYSHSCISQHCDFIFISKYSNOKNPHCVSLYKLSSP-----DDPT 71  
 DB 390 INTGKTQLSLNMGNNNSAAPSFTNFYINTSTAKPYTYILKDAQKQVKELONNDDL 449  
 QY 72 CKTKFEMATILDSAGPLDYTPPEIFPSFSTTGFTLYGMLYKPHDQPGKKYPTVLFIY 131  
 DB 450 NKLKS-----DNFIAKEFTITNAAGDONNAMIKRKNDRPAKTYVVFNFQYS 497  
 QY 132 GP----- 142  
 DB 498 GPGSQOVANSMQDNGIWFMDLAQKGYLVVCDVGRGTGRGTYKKYTKYKYLKAEIRED 557  
 QY 143 VEGLOYLASRYDIDLRVGIHGSYGYSLSMALMQRSDIPRVALAGAPYTLWIFYD 202  
 DB 558 ITRAKRLGMO-STVDSRIGIFGMSYGGFMASIAMTKGADVFMGIAVAPVTIMRYDST 616  
 QY 203 YTERYMGHPDQNEGYVLGSVAMQAEKFPSEPNRLLLHGFLENVHFAHTSILSLFLV 262  
 DB 617 YTERFLQGTQENNDQGDVNSPTVAKL--KGFLLIHGTADDNVHFNQNSMEFSALIQ 673  
 QY 263 AKRPYDLQIYPOERHSIRVPSGHEYLHLYLOENT 300

Db 674 NKKQFDMAYPDKNHSIIIGNTRPOLYKMTNTYLENL 711

## RESULT 5

T41703

dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C&gt;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C/Accession: T41703

R:Murphy, L.; Harris, D.; Bartell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1998

A:Reference number: 222011

A:Accession: T41703

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-793 &lt;NTR&gt;

A:Cross-references: EMBL:AL031180; P1DN:CAA0138.1; GSPDB:GNO0066

A:Experimental source: strain 972h-; cosmid c2B11 -chimeric

C:Genetics:

A:Gene: SPAC2B11.08

A:Map position: 1

C:Superfamily: dipeptidyl-peptidase IV

Query Match 21.4%; Score 359.5; DB 2; Length 793;  
Best Local Similarity 31.5%; Pred. No. 4,1e-23;  
Matches 113; Conservative 43; Mismatches 128; Indels 75; Gaps 17;

5 KDSPLHLLVSVYVNGEYTRLTDRGYS--HSCCTSGHCDPFISKYSNQKNPHVSLV 61

442 KDS-TERHLLVYS-LDTLEIYGITDNGEDGYSTSPSPGDFVLLVHGPDPV---WQ 495

62 KLSPPEDPTCKTEKFATILDSAGPLPDYTPPE---FSFESTTGFTLYGMLYKPHD 117

496 ELKSTDKDYCLSLDSTNSRIKQO---LSSITLPSVEYKGLTFNDIT-FNF--MERRRNP 549

118 QPGKXPTVLFYIYGPQGO-----IEIDDOVEG-----LQY 148

550 DVNKKTPVLFPAVGGPSQOVAKLFRVDFQAVLASHDPFEFIVTLDGRTGNGNAPRY 609

149 LASR-----YD-----FTLDLVGIGHGSGYGYSLMALMQRSDIFRYALA 189

610 SVSRHLGEMESYDQGAQKFMADLPFVDENHVGIMGSSYGYTLTKTL-ETQDVFSYGM 668

190 GAVVTLMIFDYDGYTERYMGHPDQNEGYLLGSVAMQAEKFPSEPRULLLHGFLENVA 249

669 VAPVDMRLYDSYTERYMDLPQYNKEG-YKNSQIHDIYERF-KQLKRFVAHGTGDVNH 726

250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

Db 727 FQSHMLMGLNLANLCVNDMAVFPDSASHI-----SVHNSLSIYRLSEWIDALGRI 781

QY 250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

A:Accession: A46465

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-20 &lt;VIV&gt;

A:Experimental source: M14.T thymoma cells, Swiss nu/nu

A:Note: sequence extracted from NCBI backbone (NCBI:P42236)

R:Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.

Biochemistry 33, 15204-15214, 1994

A:Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.

A:Reference number: A56030; MUID:95092780; PMID:7999781

A:Accession: A56030

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 746-760 &lt;BER&gt;

A:Cross-references: GB:U12620

C:Genetics:

A:Gene: CD26

C:Superfamily: dipeptidyl-peptidase IV

C:Keywords: dipeptidylpeptidase hydrolase; glycoprotein; transmembrane protein

F:213,223,315,514,679/Binding site: carboxylate (Asn) (covalent) #status predicted

F:624,702,734/Active site: Ser, Asp, His #status predicted

Query Match 20.4%; Score 342.5; DB 1; Length 760;  
Best Local Similarity 28.0%; Pred. No. 1.2e-21;  
Matches 101; Conservative 51; Mismatches 120; Indels 89; Gaps 14;

5 KDSPLHLLVSVYVNGEYTRLTDRGYS--HSCCTSGHCDPFISKYSNQKNPHVSLV 56

417 KEMPGGRNLYKI-----QLTD-HTVVKCLSDINPERCQYVAVSPSKAKYQLG 465

57 C-----SVLYKSSPEDPTCKTEKFATILDSAG---PLPDYTPPE---IFSFTTGFT 106

466 CMGPGRLPYTLHRSSTHKEKRLVLE-----DNSALDRMLQDVQMPKGLDFIVLNTRF 518

107 LYGMLYKPHDLQPGKXPTVLFYIYGP-----DLSALDRMLQDVQMPKGLDFIVLNTRF 133

519 WYGMILPFP-FDKSKYVPLLDVYAPCSQKADASRLIMATYLASTENITVASFQGRS 577

134 -----OCQIEIDDOVEGLQYLASRYDFIDLVGIGHGSGYGYSLMALMQ 179

578 GYGGDKIMHAINRRLCTLEVEDQIEAARQFV-KMGFVDSKRVAILMGWSYGYVTSWVLGS 636

180 RSDIFPVATAGAPVTLMIFDYDGYTERYMG--HPDQNEGYLLGSVAMQAEKFPSEPRULLLHGFLENVA 237

637 GSGVFCGIAVAVSSWEYDSYTERYMGRLPIPEDNLHYRNSYMSRAEHF--KOVEX 694

238 LLLHGLDENVHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESGEHELHLHYLO 297

695 LLHGTADNVHFGQSAQISKVLVDAGVDFQAMWYTDDEHGIASSTAHQHITYSHMSHFLQ 754

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

QY 298 E 298

Db 755 Q 755

OY 182 DIPRAINGAVTTLMIIPDTGYTERWYG -HPDQEGGYYCGVMAOKEPFSERNRLLL 239  
-:-|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DB 646 GVFKGGLIVAAVSRSREYDSYTEIRYMGLPRPEEDNLDRKSTWSGRBNF--KOVEYIL 703  
  
OY 240 LHGFLDENVAFTSITLSPLVRAGKPYPDLOIYPQERHSIRVPESGEHYELHLYLQE 298  
DBS IGHGTADNVHTFGQSQAISKALVDAGVDFQAMWMYTDEHDGIASSTAHQIHYSMSHFDDQ 762

RESULT 8

CDSU26  
dipeptidyl-peptidase IV (EC 3.4.14.5) - human  
N:Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)  
C:Species: Homo sapiens (man)  
C>Date: 31-Dec-1993 #sequence\_revision 23-Aug-1996 #text change 18-Jun-1999  
C:Citation: S24313; B42408; A42408; B61136; S59510; I56154; S59857; S15520  
R:Miyamori, Y.; Hayashi, Y.; Arakawa, F.; Ikebara, Y.  
B:Biochim. Biophys. Acta 1131, 333-336, 1992  
A>Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine protease involved in T cell differentiation  
A:Reference number: s24313; MUID:92329551; PMID:1352704  
A:Accession: S24313

A:Molecule type: mRNA  
A:Residues: 1-6, 'R' 8-766 <NIS>  
A:Cross-references: EMBL:X60708; NID:g35335; PIDN:CAA43118.1; PID:g35336  
R:Darmoul, D.; Lacasa, M.; Baricault, L.; Marguet, D.; Sapin, C.; Troclet, P.; Barbac, A.  
J. Biol. Chem. 267, 4824-4833, 1992  
A>Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines during cell differentiation  
A:Reference number: B42408; MUID:92165847; PMID:1347043  
A:Accession: B42408

A:Molecule type: mRNA  
A:Residues: 1-5, 'R' 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-766 <DAR1>  
A:Cross-references: GB:M60536; NID:g181569; PIDN:AAB52308.1; PID:g181570  
A:Experimental source: Intestine  
A>Note: This sequence corresponds with the author's translation  
A:Accession: A42408

A:Molecule type: mRNA  
A:Residues: 1-5, 'R' 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-711, 'G', 713-766 <DAR2>  
A:Cross-references: GB:M60536; NID:g181569  
A>Note: Sequence extracted from NCBI backbone (NCBIT:83986, NCBI:P:83988); this sequence was used as reference for the alignment  
R:Corvel, J.P.; Ferrero, A.; Chambaud, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
Castroentology 101, 618-625, 1991  
A>Title: Expression of nuclease-isomaltase and dipeptidylpeptidase IV in human small intestine  
A:Reference number: A61136; MUID:91317403; PMID:1677636  
A:Accession: B61136

A:Molecule type: Protein  
A:Residues: 1-15, 'X', 17-22 <GOR>  
R:Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.  
Biochem. J. 311, 835-843, 1995  
A>Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a 5'-proximal enhancer  
A:Reference number: S59510; MUID:96067599; PMID:7487939  
A:Accession: S59510

A:Molecule type: DNA  
A:Residues: 1-31 <BOB>  
A:Cross-references: GB:S79876; NID:g1195574; PIDN:AAB3614.1; PID:g1195575  
R:Tanaka, T.; Camarini, D.; Seed, B.; Torimoco, Y.; Dang, N.H.; Kameoka, J.; Dahlberg, U.  
Immunol. 149, 481-486, 1992  
A>Title: Cloning and functional expression of the T cell activation antigen CD26.  
A:Reference number: I56154; MUID:92325476; PMID:1352530  
A:Accession: I56154

A>Status: translated from GB/EMBL/DDBJ

A:Molecule type: mRNA  
A:Residues: 1-436, 'S', 438-766 <TAN>  
A:Cross-references: GB:W4777; NID:g180082; PIDN:AAA51943.1; PID:g180083  
R:Abdolt, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.  
Immunogenetics 40, 331-338, 1994  
A>Title: Genomic organization, exact localization, and tissue expression of the human CD26 gene  
A:Reference number: S59857; MUID:95012454; PMID:7927537  
A:Accession: S59857

A:Molecule type: DNA  
A:Residues: 1-436, 'S', 438-766 <ABB>  
A:Cross-references: EMBL:U13734

C:Genetics:  
 A:Gene: GDB:DP4  
 A:Cross-references: GDB:125239; OMIM:102720  
 A:Map position: 2q24.3-2q24.33  
 A:Introns: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3  
 C:Superfamily: dipeptidyl-peptidase IV  
 C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; proteinase; transmembr  
 F:1-6/Domain: intracellular #status predicted <INT>  
 F:7-38/Domain: transmembrane #status predicted <TM>  
 F:29-766/Domain: extracellular #status predicted <EXT>  
 F:85,92,150,219,229,281,321,520,685/Binding site: carbohydrate (Asn) (covalent) #status  
 F:630,708,740/Active site: Ser, Asp, His #status predicted

Query Match 20.1%; Score 337.5; DB 1; Length 766;  
 Best Local Similarity 28.0%; Pred. No. 3,2e-21;  
 Matches 99; Conservative 53; Mismatches 127; Indels 75; Gaps 13;

Qy 5 KDSPLEHLLVSVYVNGEVTRLTRGYSHSCCIS-QHCDPFIKYNQKNPH---C--- 57  
 Db 423 KMPGGRNLYKIQIDYKVTCL-----SCELNPERQYVSFSSKAKYQLRCGSP 475  
 Qy 58 -VSLYKSSPEDDPCKTKKEFWATILDSAGLPDTPPE---TFSESTTGTLYGMLYK 113  
 Db 476 GLPLTYLHSSVNDKGLRVLSD-NSALDKM-LQNQMPESKLDFTILNETKF-WYQMTLP 531  
 Qy 114 PHDLQPGKKYPTVLFYGGP----- 133  
 Db 532 PH-FPKSKKYPLLLVNVAGPCQKADTYPRMLMATYLAATENIYASPDGRSGYQGDKI 590  
 Qy 134 -----OGQIEIDQVEGLQYLASRYDFIDLDKRGVHWSYGYLSLWALMQRSDFRV 186  
 Db 591 MHAIRRLGTPEVEDEQIEAARQF-SKMGFVNDKRIALIMQWSGQYVTSMVLGSSGVRKC 649  
 Qy 187 AIAAGPVTLMFYDGYTERYKNG--HPDNEGGYLYGSYANQAEKFPSEPNLLILHGL 244  
 Db 650 GIAVAPVRMEYDYDVTYERVGLPTPEBDNDHYNSSTYMSRAENF--KOVEYLLIHGTA 707  
 Qy 245 DENVFATSTLSFLVRAKRPYDQIYQERHSIRVPESEGHVELHLHYLOE 298  
 Db 708 DDNVHFGQSAQSKALVDGVDFQAMWYTDDEDHGIASSIAHGHITTHSHFIKQ 761

RESULT 9  
 A49737  
 dipeptidyl aminopeptidase (EC 3.4.14.-) - Yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protease OS045; Protein YOR199c; Protein YOS50-9  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 29-Oct-1999  
 C:Accession: A49737; S45451; S60946; S67112; S71721  
 R:Santa Anna-A. S.; Herskowitz, I.  
 submitted to the Protein Sequence Database, July 1993  
 A:Reference number: A49737  
 A:Accession: A49737  
 A:Molecule type: DNA  
 A:Residues: 1-931 <SAS>  
 A:Cross-references: GB:L21944; NID:G347196; PIDN:AAA5119.1; PID:G347197  
 R:Anna-Artiola, S.S.; Herskowitz, I.  
 Yeast 10, 801-810, 1994  
 A:Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase  
 A:Reference number: S45451; MUID:95066382; PMID:7975897  
 A:Accession: S45451  
 A:Molecule type: DNA  
 A:Residues: 1-931 <ANN>  
 A:Cross-references: EMBL:L21944; NID:G347196; PIDN:AAA5119.1; PID:G347197  
 R:Galison, F.; Dujon, B.  
 submitted to the EMBL Data Library, October 1995  
 A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome  
 A:Reference number: S60938  
 A:Accession: S60938  
 A:Molecule type: DNA  
 A:Residues: 1-931 <GAL>  
 A:Cross-references: EMBL:X92441; NID:G1050762; PIDN:CAA63182.1; PID:G1050771  
 R:Boyer, J.; Fairhead, C.; Gallion, L.; Galison, F.; Michaux, G.; Thierry, A.; Dujon, B

submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S67104  
 A:Accession: S67104  
 A:Molecule type: DNA  
 A:Residues: 1-931 <BOY>  
 A:Cross-references: EMBL:275127; NID:G1420507; PIDN:CAA99437.1; PID:E252398; PID:G14205  
 A:Experimental source: strain S288C  
 R:Galison, F.; Dujon, B.  
 Yeast 12, 877-885, 1996  
 A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV  
 A:Reference number: S71721; MUID:96437977; PMID:8840505  
 A:Accession: S71721  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-931 <GAW>  
 A:Cross-references: EMBL:X92441; NID:G1050762; PIDN:CAA63182.1; PID:G1050771  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
 C:Genetics:  
 A:Gene: SGD:STE13; YC11  
 A:Cross-references: SGD:S0005745; MIPS:YOR219C  
 A:Map position: 15R  
 C:Function:  
 A:Description: involved in processing of alpha-factor prepropheromone  
 C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein  
 F:113-150/Domain: transmembrane #status predicted <TM>  
 F:377/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:785,863,886/Active site: Ser, Asp, His #status predicted

Query Match 18.7%; Score 314; DB 2; Length 931;  
 Best Local Similarity 25.6%; Pred. No. 4.6e-19;  
 Matches 100; Conservative 54; Mismatches 133; Indels 104; Gaps 12;

Qy 1 PEGKDSPLEHLLVSVYV-----NPGVETRLTRGYSHSCCISQHCDFISK 48  
 Db 562 FTANEIGVMSOHLVYSISLTDSTQNTFQSLQNP-----SKYDFYDELSSARVYISK 615  
 Qy 49 YSNQKNP-----HCVSLYKSSPEDDPCKTKKEFWATILDSAGLPDYT 92  
 Db 616 KLGPPTPKVAGPLTRVLNVAEIHDDSLQLTKDE---KFKK-----KIKYND 660  
 Qy 93 PPEIFSEST---GFTLYGLYKPHDLQPKKYPTVLFYGGPQO----- 136  
 Db 661 LP-ITSYKTVWLDQVELEINYEIKPANLNPKKPYLVNITYGPSQYFTTKSSLAPEQA 719  
 Qy 137 -----ITIDQVEG-----LOYLASRYDFIDLRV 161  
 Db 720 VVSGLDVIVLQIEPRGTGKGMSPFRWARERKLGYWEPDITEVTKFQRSQHIDESKI 779  
 Qy 162 GIHGMSYGYSLMAL-MQRSDFRVALAGAPVTLIMFYDGYTERYKNGHDPQNGQYLL 220  
 Db 780 AIMGWSYGFSLTKVELDNDGTFKYMAVAVPVNTWTLVDSVYTERYKNGQSENHGTYFE 839  
 Qy 221 GSVAMQAEKFPSEPNRLLLHGFLENVAFHTSILLSFLVRAG-KPYDLOIYQERHSI 279  
 Db 840 VSTIQNFKSPES-LKRLFIYVGTDFDDVNHQNTRLVDQMLGLITNDMHIFFPSDSHSI 898  
 Qy 280 RVPSGEHYELHLHYQENIGSRITAAIKVT 310  
 Db 899 RYHNAORIVFQKLYVWLRDAPFAERFDTEVL 929

RESULT 10  
 A30107  
 dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protease YHR028c  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 07-Jun-1990 #sequence\_revision 30-May-1997 #text\_change 23-Mar-2001  
 C:Accession: S46780; A30107  
 R:Du, Z.  
 submitted to the EMBL Data Library, June 1994  
 A:Description: The sequence of S. cerevisiae cosmid 8082.  
 A:Reference number: S46773  
 A:Accession: S46780

A: Molecule type: DNA  
A: Residues: 1-618 <DNZ>  
A: Cross-references: EMBL:U01039, NID:9500689, PIDD:AAE68879.1, PIDD:9500698, GSPDB:GN000067  
J: Roberts, C.J.; Pohlig, G.; Rothman, J.H.; Stevens, T.H.  
J: Cell Biol. 108, 1363-1373, 1989  
A: Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an invertebrate  
A: Reference number: A30107; MUID:89174971; PMID:2647766  
A: Accession: A30107  
A: Molecule type: DNA  
A: Residues: 1-82, 'H', '84-124', 'N', '126-181', 'LRLET', '189-199', 'N', '201-365', 'DFKCKERRK', 376-577  
A: Cross-references: EMBL:X15484  
A: Note: The authors translated the codon ACC for residue 572 as Asn  
C: Geneticks:  
A: Gene: SGD:DAF2, STE13, MIPS:YHR028C  
A: Cross-references: SGD:S0001070; MIPS:YHR028C  
A: Map position: 8R  
C: Superfamily: dipeptidyl-peptidase IV  
C: Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacuole  
E: 30-45/Domain: transmembrane #status predicted <TM>  
E: 63-79, 110, 139, 352, 421/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	18.5%;	Score 310;	DB 1;	Length 818;
Best Local Similarity	26.9%;	Pred. No. 8.7e-19;		
Matches	98;	Conservative	38;	Mismatches 122;
			Indels	106;
			Gaps	13;

[illegible]

## RESULT 11

hypothetical protein T23f1.7a - *Caenorhabditis elegans*  
 C.Species: *Caenorhabditis elegans*  
 C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
 C.Accession: T25173  
 R.Wilkinson, J.  
 submitted to the EMBL Data Library, October 1996  
 A.Reference number: Z19990  
 A.Accession: T25173  
 A.Status: preliminary; translated from GB/EMBL/DDBJ  
 A.Molecule type: DNA  
 A.Residues: 1-779 <MW>  
 A.Cross-references: EMBL:Z81129; PIDN:CAB03411.1; GSPDB:GN00023; CESP:T23f1.7a  
 A.Experimental source: clone T23f1  
 C.Genetics:  
 A.Gene: CESP:T23f1.7a  
 A.Map position: 5  
 A.Ions: 13/3, 52/1, 111/3, 151/2, 177/3, 346/3, 504/1, 537/1, 628/2, 683/1, 723/1

C; Superfamily: dipeptidyl-peptidase IV

Query Match	18.1%;	Score 303.5;	DB 2;	Length 779;
Best Local Similarity	25.7%;	Pred. No. 3e-18;		
Matches	93;	Conservative	62;	Mismatches 118;
				Indels 89;
				Gaps 14;

QY 7 SPLEHTLVSVNPGVTRLTDRGYSHSCIS--QHCDFFISKYSNQKNPHCVSLYKL 63

DB 431 APKPSHKSLSY-----SLADESRNSAICISCSIKNCIMAQADBDQMKIAI VSKGF 46

	..	C
	..	C
—	—	E
—		C
—		C
..		;
..	D	
..	;	
..	H	
..	C	
—	C	
—	H	
—		
—		
..		
..		
—		

**QUESTIONS**

**ANSWERS**

Db  
540 TKI STPKDTYNRKHQATPI TVHWYGGPNDONTKEATQIGFEVVASASQAATLRIDGRG 59

QY 135 -----GOIEIDDOVEGLOYLASRY-DFIDLDRVGIHGWMSYGYLSMAL 17

Db 600 SGGRGWKYRSATYGQLGTEVEDQIKAIKVLRLYRHLLDARVAVFQWSYGGFMTLSMV 65

178 MQRSD-IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNR 233

DB 660 NEAPEQFFKCAVSAPVINFAYYDAIYIERYMG--DAPLESY--SDVTKLIDNFKS--IK /

[illegible]

333

$\text{Pb}$        $\text{772}$     $\text{Pb}$     $\text{773}$   
 $\text{::}$

T25174

C/Species: *Caenorhabditis elegans*  
C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 17-Mar-2000

C/ACCESSION: 1251/4  
R:WILKINSON, J.

A;Reference number: Z19990

A;Status: preliminary; translated from GB/EMBL/DBJ

A/Residues: 1-799 <WIL>  
A/Cross-references: EMBL:Z91139, PIDN:CAP03413, CEBDB:GN0003, CESP:T323E1 7

A;Experimental Source: C10ne 123F1  
C:Genetics:

A;Map position: 5

C;Superfamily: dipeptidyl-peptidase IV

Query match	18.1%;	Score 303.5;	DB 2;	Length 799;
Best local similarity	25.7%;	Pred. NO. 3.1e-18;		

DATE	CONSERVATIVE	MODERATE	LIBERAL	DATE	CONSERVATIVE	MODERATE	LIBERAL
1951	100	100	100	1951	100	100	100
1952	100	100	100	1952	100	100	100
1953	100	100	100	1953	100	100	100
1954	100	100	100	1954	100	100	100
1955	100	100	100	1955	100	100	100
1956	100	100	100	1956	100	100	100
1957	100	100	100	1957	100	100	100
1958	100	100	100	1958	100	100	100
1959	100	100	100	1959	100	100	100
1960	100	100	100	1960	100	100	100
1961	100	100	100	1961	100	100	100
1962	100	100	100	1962	100	100	100
1963	100	100	100	1963	100	100	100
1964	100	100	100	1964	100	100	100
1965	100	100	100	1965	100	100	100
1966	100	100	100	1966	100	100	100
1967	100	100	100	1967	100	100	100
1968	100	100	100	1968	100	100	100
1969	100	100	100	1969	100	100	100
1970	100	100	100	1970	100	100	100
1971	100	100	100	1971	100	100	100
1972	100	100	100	1972	100	100	100
1973	100	100	100	1973	100	100	100
1974	100	100	100	1974	100	100	100
1975	100	100	100	1975	100	100	100
1976	100	100	100	1976	100	100	100
1977	100	100	100	1977	100	100	100
1978	100	100	100	1978	100	100	100
1979	100	100	100	1979	100	100	100
1980	100	100	100	1980	100	100	100
1981	100	100	100	1981	100	100	100
1982	100	100	100	1982	100	100	100
1983	100	100	100	1983	100	100	100
1984	100	100	100	1984	100	100	100
1985	100	100	100	1985	100	100	100
1986	100	100	100	1986	100	100	100
1987	100	100	100	1987	100	100	100
1988	100	100	100	1988	100	100	100
1989	100	100	100	1989	100	100	100
1990	100	100	100	1990	100	100	100
1991	100	100	100	1991	100	100	100
1992	100	100	100	1992	100	100	100
1993	100	100	100	1993	100	100	100
1994	100	100	100	1994	100	100	100
1995	100	100	100	1995	100	100	100
1996	100	100	100	1996	100	100	100
1997	100	100	100	1997	100	100	100
1998	100	100	100	1998	100	100	100
1999	100	100	100	1999	100	100	100
2000	100	100	100				

[illegible]

00 64 SSPED-----DPTCKTK-----FEWATII DSACBI DUYTBPEIEFESTTCEI.Y 10

Db 504 AAPHTAIVNLTRMDSDKTEHANLLYDKTYQNRRVEEAG-LPVI IKETI --KISDDFDAL 55

QY 109 GMLYKPHDLQPKKY---PTVLFIYGGPQ----- 13

Db 560 IKLSIPKDIYNRDKHQAIPLIVHVYGGPNDQNTKEATQIGIEEVVASASQAAILRIDGRG 61

QY -----GQIEIDQVEGLQILASKI-DFIDLDKRVGIHGSYGGYLSLMAL I /

```

Db      620 SGGRGWKRSAIYGQGTVEEDQIKAVLRLYHLLDARVAVFGNSYSGFMFLSNV 679
      178 MQRSD-IFRVAIAGAVTLMIFPDGTGTERYMGHPQNEQGYLGSAVMAQAEKFPSEPNR 236
      Db      680 NPAPEQFFCAVSAVAIVTFAVYDATYTERYMG--DAPLESY--SDVTKLDNFKS--TR 733
      Qy      237 LLLHGFLENVFAHTSILSLFVRAKQPYDQIYPOERHSIRVPESGEYELHLHYL 296
      Db      734 LLLHGLLDNNVFNQSAIILIDELQNRGVDFDLNVPNQAHSLSRST--HVGKMTTHL 791
      Qy      297 QE 298
      Db      792 RQ 793

```

## RESULT 13

```

119514
hypothetical protein C27C12.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T19514
R:Thomas, K.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19134
A:Accession: T19514
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-829 <MIL>
A:Cross-references: EMBL:Z69883; PIDN:CAA93743.1; GSPDB:GN00028; CESP:C27C12.7
A:Experimental source: clone C27C12
C:Genetics:
A:Gene: CESP:C27C12.7
A:Map position: X
A:Introns: 4/3; 51/1; 123/3; 166/3; 188/2; 279/2; 392/3; 501/3; 553/1; 583/1; 606/2; 647
C:Superfamily: dipeptidyl-peptidase IV

```

```

Query Match      16.9%; Score 284.5; DB 2; Length 829;
Best Local Similarity 26.1%; Pred. No. 1.5e-16;
Matches 87; Conservative 50; Mismatches 103; Indels 93; Gaps 10;

Qy      49 YSNQKNPHCVSLYK-----LSSPDDPT-----CK-----TKEFWATILDSA- 85
      497 YSNKK-----TLQKVTQDFEDCKTLGSGADPTGQRIVQCEKPFNDTRLVLDVVDTK 552
      Qy      86 -----GRLPDYTPREIF-SPESTGTFLYGMKYKHNDIQPKKYPTVLFYGGPG- 134
      Db      553 KIMLEGTKAVIPFDVPMKKFGKLKPSGIDGHYMWLTPANLLDGAKIPLLDIYGGPDS 612
      Qy      135 -----GQIEIDVOYEGLOY 148
      Db      613 KQVFKCTPTAAHIOIVSOVDIAYRIDVAGTGRGMDVKEAVYRKLGDAEVVDITLDMIRA 672
      Qy      149 LASRYDFIDLRVGHGMSYGYL-SLMAIMQRSDIFRVAIAGAPVTLMIFYDTGTERY 207
      Db      673 FINTGFIDEDIAWMSYGGFTLSKIAIKDQGEIVKCAISIAVTPDKYDSATERY 732
      Qy      208 MGHPNQEGYVYLGSAVMAQAEKFPSEPNRLLLHGFLENVFAHTSILSLFVRAKQPY 267
      Db      733 LGQPEENIQGYINTVIVIPHARNTV--KYLHAGEBDNNVHYONSARMSALQONGIHF 790
      Qy      268 DLQIYPOERHSIRVPESGEYELHLHYLOENL 300
      Db      791 TQLVYANEAHSL-----SHKLPHLYGGEVQRF 817

```

## RESULT 14

```

A11793
dipeptidyl aminopeptidase-like protein (BC 3.4.14.-) short form, DPPX-S - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A11793
R:Mada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthhold, R.J.; Shimaaki, S.

```

```

Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A:Title: Differential expression of two distinct forms of mRNA encoding members of a di
A:Reference number: A41793; MID:92108018; PMID:1729689
A:Accession: A41793
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-803 <MAD>
A:Cross-references: GB:M76429; NID:g408719; PIDN:AA041623.1; PID:g408720
A>Note: Sequence extracted from NCBI backbone (NCBI:P:75138)
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; glycoprotein
F:257,342/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match      15.6%; Score 262; DB 2; Length 803;
Best Local Similarity 25.0%; Pred. No. 1.3e-14;
Matches 89; Conservative 45; Mismatches 152; Indels 70; Gaps 9;

```

```

Qy      1 FEETKDSPLENHLVYVSVNPEVTR-----LTDRGYSHCCISQHCDFISKYSQKN 54
      Db      440 FLSTEDLPRRRQLYASIV--GSFNRQCLSCDLVNCYTFSAFSFGADPFLKCGPGV 497
      Qy      55 PRCVSLYKLSPEDDPTCKTKF-FWATILDSAGPLDYTPREIFSESTGTFLYGMLYK 113
      Db      498 P-TVSVHNTTDKKMFPLETNEHVQKASDRQMPKVEYAKIE-----TDDVNLPIQILK 550
      Qy      114 PHDLQPKKYPVNLFIYGGP----- 133
      Db      551 PATFTDPAHYPLLVLDGTPGSGSAEKAFAVETVMSSHGAVVVCKDGRSGFGQTRL 610
      Qy      134 -----QGIEIDDDQVEGLQYLAIRYDFIDLRVGHGMSYGYLSLMAIMQRSD----- 182
      Db      611 LHEVRRRLGSLSEKQMEAVRYML-KEPTIDKTRVAVFCKDYGGIYSTLLPAKGGQAP 669
      Qy      183 IFRVAIAGAPVTLMIFYDTGTERYMGHPDQNEQGYLGSAVMAQAEKFPSEPNRLLLHG 242
      Db      670 VPSGCSALSPIDPFKLYASAFSERYLGLHLDNRAVEMAKVAHRSAL--EGQFLVIVA 727
      Qy      243 FIDENVHFAHTSILSLFVRAKQPYDQIYPOERHSIRVPESGEYELHLHYLOE 298
      Db      728 TADEKIHQVETALITQLIKGRANYSLOIYPESHYFSSALQOHLHRSILGFVE 783

```

## RESULT 15

```

168600
dipeptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C:Accession: 168600
R:Yokotani, N.; Doi, K.; Wenthhold, R.J.; Mada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-rela
A:Reference number: 154331; MUID:93172805; PMID:8103397
A:Accession: 168600
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-803 <RES>
A:Cross-references: GB:M6860; NID:g306707; PIDN:AAA35761.1; PID:g306708
C:Superfamily: dipeptidyl-peptidase IV

```

```

Query Match      14.8%; Score 248; DB 2; Length 803;
Best Local Similarity 23.0%; Pred. No. 2.1e-13;
Matches 84; Conservative 53; Mismatches 139; Indels 90; Gaps 10;

Qy      1 FEETKDSPLENHLVYVSVNPEVTRLTDRGYSH--SCCISQHCDFISKYSQKNPHC 57
      Db      440 FLSTEDLPRRRQLYASN-----TEGNFNRQCLSCDLVNCYTFSAFSHMD--- 486
      Qy      58 VSLYKLSPP-----EDPTCKTKF-----FWATILDSAGPLDYTPREIFSESTT 103
      Db      487 FPLKCEGPGVPMVTHNTTKKMFPLETNEHVAKKAIINDRQMPKVEYADIEI-----D 540
      Qy      104 GFTLYGMLYKPHDLQPKKYPVNLFIYGGPGQ----- 136

```

```

Db      541 DYNLPWQILKPATFTDTTHYPPLLVDGTPGSQVAEKEFEVSWETVMVSSHGAVVVKCDG 600
Qy      137 -----IEIDQVEGLQYLASRYDFIDLDRVGIHGMVGYLSLMA 176
Db      601 RSGSGOCTKLHEVRRRLGLLEKDOMEAVRTML-KEQYIDRTRVAVFGDYGGYSTYI 659
Qy      177 LMQRSD----IFRVAIAGAPVTLMIFYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPs 232
Db      660 LPAKGENQOCTFTCSALSPTIDFKLYASAFSERIYGLHGLDNRAVEMTKVAHRVSAL-- 717
Qy      233 EPNRLILLHGFUDENVHFAHTSITLSFLVPAGKPYDLOIYPOERHSIRVPESGEHYELHL 292
Db      718 EEQOFLIHPTADEKIHFOHTAELITOLIRKANYSLOIYPDESHYFTSSSLKQHLYRSI 777
Qy      293 LHYLOE 298
Db      778 INFVE 783

```

Search completed: October 15, 2003, 17:14:46  
 Job time : 9.14576 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 (Search time 4.9182 Seconds)

(without alignments)  
2984.152 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1680

Sequence: 1 FEGTSDSPLEHLYVSYVN.....HLHYLGKSLAIAKVI 310

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.5	21.4	793	1	YEAB_SCHPO
2	345.5	20.6	765	1	DPB4_BOVIN
3	344.5	20.5	767	1	DPB4_RAT
4	341.5	20.3	760	1	DPB4_MOUSE
5	335.5	20.0	765	1	DPB4_FELCA
6	333.5	19.9	766	1	DPB4_HUMAN
7	314	18.7	931	1	STI3_YEAST
8	310	18.5	818	1	DAF2_YEAST
9	309.5	18.4	760	1	SEPR_MOUSE
10	302	18.0	761	1	SEPR_HUMAN
11	262	15.6	863	1	DPB6_BOVIN
12	248	14.8	865	1	DPB6_HUMAN
13	244	14.5	853	1	YDZP_SCHPO
14	241	14.3	804	1	DPPE_MOUSE
15	240	14.3	859	1	DPPE_RAT
16	152.5	9.1	657	1	YUXL_BACSU
17	140	8.3	732	1	ACPH_HUMAN
18	138	8.2	732	1	ACPH_RAT
19	126.5	7.5	732	1	ACPH_PIG
20	112.5	6.7	740	1	YL31_CHEEL
21	104.5	6.2	298	1	Y812_ARCFU
22	103.5	6.2	689	1	PPCB_AERYH
23	94	5.6	722	1	Y174_RICPR
24	92.5	5.5	4451	1	GRSB_BACBR
25	91	5.4	484	1	PICB_HUMAN
26	89.5	5.3	1121	1	EXSC_HAEIN
27	88	5.2	1181	1	ITAZ_HUMAN
28	87.5	5.2	276	1	TODF_PSEPU
29	87	5.2	299	1	YUGB_YEAST
30	87	5.2	3587	1	TYCB_BREPA
31	86.5	5.1	264	1	COMT_RAT
32	86.5	5.1	403	1	KAPR_BIAEM
33	86.5	5.1	543	1	FAT2_YEAST

34	86	5.1	352	1	PSBD_SPIOL
35	86	5.1	353	1	PSBD_ARATH
36	86	5.1	353	1	PSBD_PINTH
37	86	5.1	353	1	PSBD_POPDE
38	85.5	5.1	286	1	DHA2_MYCTU
39	85.5	5.1	371	1	YM18_PSEAE
40	85	5.1	156	1	RNH_ZYMMO
41	84	5.0	493	1	PCD_ARROX
42	83.5	5.0	353	1	PSBD_WHEAT
43	83.5	5.0	705	1	PPCE_FLAME
44	83.5	5.0	705	1	PPCF_FLAME
45	83	4.9	739	1	BAC1_MOUSE

## ALIGNMENTS

```

RESULT 1
YEAB_SCHPO
ID 014073,
AC 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative dipeptidyl aminopeptidase C2E11.08 (EC 3.4.14.-).
GN SPACNKA.08 OR SPACZEL1.08
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1] 9
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Baeham D., Bowman C.M.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Godle A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purrelle B.,
RA Goffeau A., Cadiu E., Dreano S., Gloux S., Lelure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;
RL The genome sequence of Schizosaccharomyces pombe.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
vacuoles (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

```

DR	EMBL; AL031180; CAA20318.1; -	
DR	PIR; T41703; T41703.	
DR	GeneDB; SPombe; SPACUNC4_08; -	
DR	InterPro; IPR002469; Dppiv_N term.	
DR	InterPro; IPR001375; Peptidase_S9.	
DR	InterPro; IPR002471; Prol_endopep_ser.	
DR	InterPro; IPR000379; Ser_estr_site.	
DR	Pfam; PF00930; Dppiv_N_term; 1.	
DR	Pfam; PF00326; Peptidase_S9; 1.	
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.	
KW	Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;	
KW	Serine protease; Transmembrane; Glycoprotein; Signal anchor.	
FT	DOMAIN	1 24
FT	TRANSMEM	25 45
FT		
FT	DOMAIN	46 793
FT	ACT SITE	647 647
FT	ACT_SITE	722 722
FT	ACT_SITE	755 755
FT	CARBOHYD	101 101
FT	CARBOHYD	136 136
FT	CARBOHYD	246 246
FT	CARBOHYD	299 299
FT	CARBOHYD	303 303
FT	CARBOHYD	324 324
FT	CARBOHYD	336 336
FT	CARBOHYD	377 377
FT	CARBOHYD	384 384
FT	CARBOHYD	407 407
FT	CARBOHYD	535 535
FT	CARBOHYD	761 761
FO	SEQUENCE	793 AA; 91304 MW; 20B70F97FE231463 CRC64;

	Query Match Similarity	21.4%;	Score 359.5;	DB 1;	Length 793;	
	Best Local Similarity	31.5%;	Pred. No. 4,2e-24;			
	Matches	113;	Conservative	43;	Mismatches 128;	Indels 75; Gaps 17;;
Qy	5	KDSPLEHLLVYVSYNPESEVTRLDGRYS--HSCCISQHCDFISKYSNOKNPCHVCSLY	61			
Db	442	KDS-TERHLVVYS-IDTLEIYGITDNGDEGGYSTSPFSDPFYLANHGCDVP---WQ	495			
Oy	62	KLSSPEDPPTCKTEKFMTILDSAGELPDYTPREL----SFSESTGTFTLGLMKPHDL	117			
Db	496	ELRSTKDKDYCLSLSETNSRLAQO--LSSTILPSVEYEGKLTFNDDTT-PNF-MERRPNF	549			
Oy	118	QPKKKYPTVLPIYGGPOGQ-----IEIDQVEG-----LOY	148			
Db	550	DVNKKYEVLPFAYGCGPGQQVAKLPRVDFOAYLAHPDEFIVTLDDRGTFGNMGRFY	609			
Oy	149	LASR-----YD-----FIIDLVRGHGMSYGGYLISMALMQSDIFRAVALA	189			
Db	610	SVSRIHGWESYDGOQAQGFMADLPFDVENHVIGWSYGGYLTLLKTL-ETQDVFSYGM	668			
Oy	190	GAPVTLMIFYDTGYTERMYMGHPDQOGQYUUGSVAMAQKEPSEPNNRLLTHGFLENVH	249			
Db	669	VAPTDMRLYDSVYTERYMDELPOYNKEG-YKNSQIHDEYE-KOLKRPFVAHGCDNVH	726			
Oy	250	FATSIILSFLVRAG-KPYDLQIYPQERHSIRVPESGEHYELLHLHYLOENIGSRIAAL	307			
Db	727	FOHSMHMLDGINTLNCYVMADVAFPPDSANSHI-----SYHNASTLIYHRLSIEWIGDALGRI	781			
<b>RESULT 2</b>						
DPp4_BOVIN	ID_DPp4_BOVIN	STANDARD,	PRT,	765	AA.	
AC	P814Z5; O8MMG8;					
DT	15-DEC-1998 (Rel. 37, Created)					
DT	28-FEB-2003 (Rel. 41, Last sequence update)					
DT	28-FEB-2003 (Rel. 41, Last annotation update)					
DE	Dipeptidyl peptidase IV (EC 3.4.14.5) (DBP IV) (T-cell activation					
DE	antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)					
DE	(Activation molecule 3) (ACT3) (WC10).					
NN	DPp4 OR CD26.					

OS Bos taurus (Bovine).  
OC Eumetazoa, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymphocytes;  
RX MEDLINE=22067734; PubMed=12073152;  
RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,  
RT Bohnach G.A.;  
RT "Molecular characterization of bovine CD26 upregulated by a  
RL streptococcal superantigen.";  
RT Immunogenetics 54:216-220(2002).  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.  
RC TISSUE=Thymus;  
RX MEDLINE=22021197; PubMed=11981836;  
RA Gliddon D.R., Howard C.J.;  
RT "CD26 is expressed on a restricted subpopulation of dendritic cells in  
RT vivo";  
RL Eur. J. Immunol. 32:1472-1481(2002).  
RN [3]  
RP SEQUENCE OF 1-24.  
RC TISSUE=T-cell;  
RX MEDLINE=21482004; PubMed=11598101;  
RA Lee S.-U., Petersen W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,  
RT Naessens J., Bohnach G.A.;  
RT "Identity of activation molecule 3 on superantigen-stimulated bovine  
RL cells is CD26";  
RL Infect. Immun. 69:7190-7193(2001).  
RN [4]  
RP SEQUENCE OF 537--546.  
RC TISSUE=Kidney;  
RX MEDLINE=98293306; PubMed=9629661;  
RA Ben-Shoshan I., Parola A.H.;  
RT "The CP-1 subunit of adenosine deaminase complexing protein from calf  
RT kidney is identical to human, mouse, and rat dipeptidyl peptidase  
RV IV.";  
RL Comp. Biochem. Physiol. 119B:289-293(1998).  
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from  
CC polypeptides having unsubstituted N-termini provided that the  
CC penultimate residue is proline. Binds and regulates the activity  
CC of ADA.  
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb[-  
CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided  
CC Xcc is neither Pro nor hydroxyproline.  
CC -1- SUBUNIT: Homodimer.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in  
CC a soluble form (By similarity).  
CC -1- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and  
CC several immune system tissues.  
CC -1- PFM: The soluble form (SDP) derives from the membrane form (MDPP)  
CC by proteolytic processing (By similarity).  
CC -1- SIMILARITY: Belongs to peptidase family S9B.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isdb-sib.ch/announce/>  
CC or send an email to [license@isdb-sib.ch](mailto:license@isdb-sib.ch)).  
CC -----  
DR EMBL, AF461806; AAL67836.1; -;  
DR EMBL, AY056834; AAL23628.1; -;  
DR MEROPS, S09\_003; -;  
DR InterPro, IPRO02471; Prol endopep ser.  
DR InterPro, IPRO00379; Ser esters site.  
DR Pfam, PF009930; DPPIV\_N term; 1.  
DR Pfam, PF00326; Peptidase\_S9; 1.  
DR PROSITE, PS00708; PRO\_ENDOPEP\_SER; 1.  
DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;

KM Transmembrane; Glycoprotein; Signal-anchor.  
 FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM  
 FT CHAIN 38 765 (MDPP).  
 FT CHAIN 38 765 (DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM  
 FT DOMAIN 1 6 (SDPP) (BY SIMILARITY).  
 FT TRANSMEM 7 29 CYTOPLASMIC (POTENTIAL).  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).  
 FT ACT SITE 629 765 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 84 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 765 AA; 88369 MW; E32165421F43E116 CRC64;

Query Match 20.6%; Score 345.5; DB 1; Length 765;  
 Best Local Similarity 28.9%; Pred. No. 7.1e-23;  
 Matches 101; Conservative 46; Mismatches 136; Indels 67; Gaps 10;

QY 5 KDSPLEHLVYVSVNPEVETRLT-----DRGVSHSCCISQHCDFISKYSNOMKPHCVS 59  
 422 KGMFGANLVKRIQNDYTKVTCSCELNPRCQYVSFSQBAKYVGLRCSGPELP----- 477  
 DB 60 LYKLSSEDDPTCKTKFPAFTILDSAGPLPD-YTPPEIFSEFSTGTTL-YGMLYKHDL 117  
 478 LYLTHSNMNNKEXELRVLE---NNSDLQVLDVQVMPKSLFHLHGKTFYQMILPPL-F 533  
 QY 118 QPKKYPVTFIYGGP----- 133  
 DB 534 DKSKKPLLEVAAGPSCQADAFRLNMTATYSTNIIVASFDGSGSYQGDKNHAI 593  
 134 ---QGQLEIDDQVEGLQYLARVDFIDLDRVGHGWSYGYLSMALMORSDFRVAIAG 190  
 DB 594 NRRLGTEVEDQLEATRQF-SKMGFVDDKRIAIKMGWYGVGVSMVLGAASGVFKCIIAV 652  
 QY 191 APYTLTFYDTGYTTERYWG--HPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGFLEDEV 248  
 653 APVSKMEYDVSYTERYMGLPTEPDNLSYRNSITVMSRAENF--KQVEYLLIHGTADDNV 710  
 QY 249 HFAMTSLLSFLVRAGKPYDLQIYPOBRSIRVSESEHMLHLHTLOE 298  
 DB 711 HFQOSAOISKALYDAGVDPOSMWTTDEDHGLASTAHQHIYTHMSHFLKQ 760

RESULT 3  
 DDP4 RAT STANDARD; PRT; 767 AA.  
 AC P14730;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) ('T-cell activation  
 DE antiserum CD26) (GPI10 glycoprotein) (Bile canalicular domain-specific  
 DE membrane glycoprotein).  
 GN DDP4 OR CD26.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=89123496; PubMed=2563382;  
 RA Ogata S., Miumi Y., Ikehara Y.;

RT "Primary structure of rat liver dipeptidyl peptidase IV deduced from  
 its cDNA and identification of the NH2-terminal signal sequence as  
 the membrane-anchoring domain.";  
 RT J. Biol. Chem. 264:3596-3601(1989).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88065516; PubMed=3479775;  
 RA Hong W., Doyle D.;  
 RT "cDNA cloning for a bile canalicular domain-specific membrane  
 glycoprotein of rat hepatocytes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).  
 RN [3]  
 RP SIGNAL-ANCHOR.  
 RX MEDLINE=90338089; PubMed=1974258;  
 RA Hong W., Doyle D.;  
 RT "Molecular dissection of the NH2-terminal signal/anchor sequence of  
 rat dipeptidyl peptidase IV.";  
 RL J. Cell Biol. 111:323-326(1990).  
 RN [4]  
 RP SEQUENCE OF 281-302.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94128339; PubMed=7905271;  
 RA Iwaki-Egawa S., Watanabe Y., Fujimoto Y.;  
 RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney  
 dipeptidyl peptidase IV.";  
 RL Biol. Chem. Hoppe-seyler 374:973-975(1993).  
 CC -1- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM  
 CC POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE  
 CC PENULTIMATE RESIDUE IS PROLINE.  
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-  
 CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided  
 CC Xcc is neither Pro nor hydroxyproline.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN  
 CC A SOLUBLE FORM.  
 CC -1- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)  
 CC BY PROTEOLYTIC PROCESSING.  
 CC -1- SIMILARITY: Belongs to peptidase family S9B.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; J04591; AAA41096.1; -;  
 DR EMBL; J02997; AAA41272.1; -;  
 DR PIR; A39914; A39914.  
 DR MEROPS; S09.003; -;  
 DR InterPro; IPR002469; DPPIV\_N\_term.  
 DR InterPro; IPR001375; Peptidase\_S9.  
 DR InterPro; IPR002471; Prol\_endopep\_ser.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 DR Pfam; PF00930; DPPIV\_N\_term; 1.  
 DR Pfam; PF00326; Peptidase\_S9; 1.  
 DR PROSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
 KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;  
 KM Transmembrane; Glycoprotein; Signal-anchor.  
 FT CHAIN 1 767 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM  
 FT CHAIN 37 767 (MDPP).  
 FT CHAIN 37 767 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM  
 FT CHAIN 281 767 (SDPP).  
 FT DOMAIN 1 6 DIPEPTIDYL PEPTIDASE IV 60 KDA SOLUBLE  
 FT TRANSMEM 7 28 FORM.  
 FT CYTOPLASMIC (POTENTIAL).  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 29 767 EXTRACELLULAR (POTENTIAL).  
 FT ACT SITE 631 631 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 709 709 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 741 741 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 686 686 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 38 38 A -> R (IN REF. 2).
FT CONFLICT 183 183 I -> T (IN REF. 2).
FT CONFLICT 332 332 T -> N (IN REF. 2).
FT CONFLICT 352 352 C -> V (IN REF. 2).
FT CONFLICT 394 394 V -> D (IN REF. 2).
FT CONFLICT 562 562 L -> F (IN REF. 2).
FT CONFLICT 624 624 R -> Q (IN REF. 2).
SQ SEQUENCE 767 AA; 88003 MM; AA170CECGF084652 CMC64;

Query Match 20.5%; Score 344.5; DB 1; Length 767;
Best Local Similarity 29.0%; Fred. No. 8.7e-23;
Matches 104; Conservative 50; Mismatches 120; Indels 85; Gaps 15;

QY KDSPLEHLLVYVYNPGEVRLTRGYSHSCIS-----QHCPFF--ISKYGNQKNPH 56
Db KEMFGCNLYKI-----QLTD--HTNKKCSCLNPERCQYYSLSKEAKYQLG 472
QY C---VSLYKLSPEDEPTCKTKFMAITIDAGPLPDYPPPE--IFSFGTGTLYG 109
Db CRGGLPLTYLHRSTQKEKRLVED--NSALDKM--LQGVQPSKLDPIVLETRF-WYQ 528
QY 110 MLYRPHLOPCKKRYPTLYFYGP----- 133
Db 529 MLRPPH-FDKSKRYPLLDVAGPQSOQADAFRLNATYASTENITIVASFDGSGCYQ 587
QY 134 -----QCGIEIDQVEGL-QYLASHRYDFIDDRVGHGMSGYSLMALMORS 181
Db 588 GDKIMHAINKRLGLFVEDQLEAARQFL--KMGFVDSKRAVIMKMSIGYITSNVLGSGS 645
QY 182 DIFRAVIAAGAPVTIWTIFGYTERYWG--HPDNGOGYVGVAMQAEKFPSPNRLLT 239
Db 646 GVFCGIAVAVPVSWEYDSVYTERYMG.LPPEENLDLHRYSTVMSRAENE--KQVEYLL 703
QY 240 LHGFLDNVPHAFHNSITLSPYRAGKPYDLDIYQOEHRHSIRVPSGGEYELHLLAYOE 298
Db 704 IHGTRADNVHFOQSAQISKALVDAGVDFQAMWYDEDEHGASSTAHQHTYSHMSHPLOQ 762

RESULT 4
DPP4_MOUSE
ID DPP4_MOUSE STANDARD; PRT: 760 AA.
AC P28842;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
antigen CD26) (thymocyte-activating molecule) (THAM).
GN DPP4 OR CD26.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Thymus;
RX MEDLINE=92129286; PubMed=1370813;
RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
Pierres M.;
RT "cDNA cloning for mouse thymocyte-activating molecule. A
multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
subgroup of serine proteases."
RL J. Biol. Chem. 267:2200-2208 (1992).
RN [2]
RP REVISIONS.
RA Marguet D.A.;

```

```

RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BiO.A; TISSUE=Liver;
RX MEDLINE=95092780; PubMed=7999781;
RA Bernard A.-M., Matei M.-G., Pierres M., Marguet D.;
RL "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
Biochemistry 33:15204-15214 (1994).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,
RA Datchenko L., Mausina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Farey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RP SEQUENCE OF 1-20.
RX MEDLINE=91302787; PubMed=1712807;
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
RA Bernard A.-M., Gorvel J.-P., Pierres M.;
RT "Evidence that thymocyte-activating molecule is mouse CD26
(dipeptidyl peptidase IV).";
RL J. Immunol. 147:447-454 (1991).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
polypeptides having unsubstituted N-termini provided that the
penultimate residue is proline.
CC CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
a soluble form.
CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
by proteolytic processing.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements
or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X58384; CA041274.1; -
DR EMBL; U12620; AAA82213.1; -
DR EMBL; U12599; AAA82213.1; JOINED.
DR EMBL; U12600; AAA82213.1; JOINED.
DR EMBL; U12601; AAA82213.1; JOINED.
DR EMBL; U12602; AAA82213.1; JOINED.
DR EMBL; U12603; AAA82213.1; JOINED.
DR EMBL; U12604; AAA82213.1; JOINED.
DR EMBL; U12605; AAA82213.1; JOINED.
DR EMBL; U12606; AAA82213.1; JOINED.
DR EMBL; U12607; AAA82213.1; JOINED.
DR EMBL; U12608; AAA82213.1; JOINED.
DR EMBL; U12609; AAA82213.1; JOINED.

```

```

DR EMBL; U12610; AAA82213.1; JOINED.
DR EMBL; U12611; AAA82213.1; JOINED.
DR EMBL; U12612; AAA82213.1; JOINED.
DR EMBL; U12613; AAA82213.1; JOINED.
DR EMBL; U12614; AAA82213.1; JOINED.
DR EMBL; U12615; AAA82213.1; JOINED.
DR EMBL; U12616; AAA82213.1; JOINED.
DR EMBL; U12617; AAA82213.1; JOINED.
DR EMBL; U12618; AAA82213.1; JOINED.
DR EMBL; U12619; AAA82213.1; JOINED.
DR EMBL; BC022183; AAA82213.1; JOINED.
DR MEROPS; S09.003; -.
DR MGD; MGI:94919; DPP4.
DR InterPro; IPR002469; DPP4V N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_ester_site.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
DR Hydroxylase; Aminoacylase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 760
FT CHAIN 1 760
FT CHAIN 37 760
FT CHAIN 1 6
FT TRANSMEM 7 28
FT DOMAIN 29 760
FT ACT_SITE 295 466
FT ACT_SITE 624 624
FT ACT_SITE 702 702
FT ACT_SITE 734 734
FT CARBOHYD 83 83
FT CARBOHYD 90 90
FT CARBOHYD 113 113
FT CARBOHYD 213 213
FT CARBOHYD 223 223
FT CARBOHYD 315 315
FT CARBOHYD 328 328
FT CARBOHYD 514 514
FT CARBOHYD 679 679
SQ SEQUENCE 760 AA; 87436 MW; A5F644B46E4A3DF8 CRC64;

Query Match 20.3%; Score 341.5; DB 1; Length 760;
Best Local Similarity 28.0%; Pred. No. 1.6e-22;
Matches 101; Conservative 51; Mismatches 120; Indels 89; Gaps 14;

5 KDSPLHHLVYVSVNPGEVRLTDRGYSHSCIS-----QHCDPFIKYNOKNPH--- 56
417 KEMPGGRNLYKI-----QLTD--HNNVKCLSCDLNPERCQYAVSFSEAKYQLG 465
57 C---VSLYKSSPEDDPTCKTEKEMATILDSAG---PLPDYTPPE---IFSEESTTGT 106
466 CMGPGLEPLTLHRSSTDHKEKRLYLE-----DNSALDRMDQVMPKSKLDIFLNLNTRF- 518
107 LYGMLYKHPDLQPKKKYPTVLFYIGP----- 133
519 WQOMLTPPH-FPKSKYPLLLDLYVAPCSQKADASRLNWTATLASTENIIVASFDGRGS 577
134 -----OQOIEIDDOVEGLQYLASRYDFIDLDRAVGHGMSYGYLSLMAQLQ 179
578 GYGQDKIMHAINRRLGTLLEVEDQIEAARQFV-KMGVVDKRAVAIKMSYGYVTSNVLGS 636
DB 180 RSDIRVALAGAPVTLMTFYDTGYTRRYNG--HPDNEGGYLYGSVAAQAEKPPSEPNRL 237
DB 637 GSGVFCGGAIVAPVSMETYSVTERYVGLPLPEBNLDHYNNSTWMSRAEHF--KQVEY 694
QY 238 LLLHGFLENDVFAHTSILSLFVNRAGKPYDQLQYPOEHSRIVRPESGHHYELHLHYQL 297
DB 695 LLIHGTADNVHFOQSAQISKALVDAGVDFOAMWYTDDEHGIASTAHQIHYHSHFLLQ 754

```

```

QY 298 E 298
DB 755 Q 755

RESULT 5
DPP4_FELCA STANDARD; PRT; 765 AA.
ID DPP4_FELCA
AC 09N217;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DPP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxId=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue."
RL Immunogenetics 50:366-368(1999).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC penultimate residue having unsubstituted N-termini provided that the
CC polypeptide residue is proline (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family 59B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB023954; BAA92344.1; -.
DR MEROPS; S09.003; -.
DR InterPro; IPR002469; DPP4V N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_ester_site.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW Hydroxylase; Aminoacylase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765
FT CHAIN 38 765
FT DOMAIN 1 6
FT TRANSMEM 7 29
FT ACT_SITE 629 765
FT ACT_SITE 707 707
FT ACT_SITE 739 739
FT CARBOHYD 84 84
FT CARBOHYD 91 91

```

FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 765 AA; 88213 MW; 3EFC98A22B175D9 CRC64;

Query Match 20.0%; Score 335.5; DB 1; Length 765;  
 Best Local Similarity 27.5%; Pred. No. 5,5e-22;  
 Matches 98; Conservative 53; Mismatches 128; Indels 77; Gaps 14;

12 HLTVVNVNPG-----EVRLLTRGVSNSCCIS-----QHCDPEFISYNSQKPH---C- 57  
 413 YLYTINSEYKMGNGNLYKIQLNDYTKVACLSCELPKRCQYYSFSREAKYQLRCS 472  
 58 ---VSLYKLSPPDDPTCKTKFEWATLDSAGPLPDYTPPE---IFSFESTTGFTLYGML 111  
 473 GPGLPLTYLHRSNDELRLVLED--NSALDGM--LQEVQWPSKKLDFTILNETKF--WYQMT 528  
 112 YKHNDLQPKKRYTVLEFYGSP----- 133  
 529 LPPH-FDTSKRYPLLIDYVAGPSCQKADALFRLNMTATYSTENIIVASFDRGSGYOGD 587  
 134 -----QGQIEIDDOVEGLQYLARSYDFIDLRVGHGMSYGYLSLMLMQRSDIF 184  
 588 KIMHAYNRRLGTEVEVDQIEAARQF-SKMGFVUDKILALMGWYGYGYSMTVGAAGSGVF 646  
 185 RVAIAGAPVTLWIFDTGYTERYMGHP-DONEGQYVIGSVAM-QAEKPEPSEPRLLLLG 242  
 647 KCGIAVAPSRMWEYDSVTERYMGLPDPDNDLYKNSITVMSRAENF--KQVEYLLING 704  
 243 FLDENHFAHTSLLSLFLVAGKPYDLQIYPOERHSIRVSESEHNEHLNHTLOE 298  
 705 TADNVNHFQOASQISKALVDAGVDFQAMWYTDDEHGIASGPAHQHITYTHMSHPKQ 760

DB 705 TADNVNHFQOASQISKALVDAGVDFQAMWYTDDEHGIASGPAHQHITYTHMSHPKQ 760

QY 243 FLDENHFAHTSLLSLFLVAGKPYDLQIYPOERHSIRVSESEHNEHLNHTLOE 298

DB 705 TADNVNHFQOASQISKALVDAGVDFQAMWYTDDEHGIASGPAHQHITYTHMSHPKQ 760

RESULT 6  
 DPP4 HUMAN STANDARD; PRT; 766 AA.  
 ID DPP4\_HUMAN STANDARD; PRT; 766 AA.  
 AC P27487;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (TPI03) (Adenosine deaminase complexing protein-2) (ADAP).  
 GN DPP4 OR ADCP2 OR CD26.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=92329551; PubMed=1352704;  
 RA Mismuni Y., Hayashi Y., Arakawa F., Ikehara Y.;  
 RT "Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine proteinase on the cell surface.";  
 RL Biochim. Biophys. Acta 1131:333-336(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=95012454; PubMed=7927537;  
 RA Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;  
 RT "Genomic organization, exact localization, and tissue expression of the human CD26 (dipeptidyl peptidase IV) gene.";  
 RL Immunogenetics 40:331-338(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Peripheral blood;  
 RX MEDLINE=92325476; PubMed=1352530;  
 RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,  
 RT Dahlberg H.N., Schlossman S.F., Morimoto C.;  
 RT "Cloning and functional expression of the T cell activation antigen CD26.";  
 RL J. Immunol. 149:481-486(1992).  
 RN [4]  
 RP ERRATUM.  
 RX MEDLINE=93171637; PubMed=8094732;  
 RA Tanaka T.;  
 RL J. Immunol. 150:2090-2090(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedl T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallary S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska D.E.,  
 RA Schenker A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 1-551 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=92165847; PubMed=1347043;  
 RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,  
 RA Troclet P., Barbat A.;  
 RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines HT-29 and Caco-2. Cloning of the complete human coding sequence and changes of dipeptidyl peptidase IV mRNA levels during cell differentiation.";  
 RL J. Biol. Chem. 267:4824-4833(1992).  
 RN [7]  
 RP SEQUENCE OF 545-766 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=91024044; PubMed=1977364;  
 RA Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;  
 RT "Isolation of a cDNA probe for the human intestinal dipeptidylpeptidase IV and assignment of the gene locus DPP4 to chromosome 2.";  
 RL Ann. Hum. Genet. 54:191-197(1990).  
 RN [8]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RX MEDLINE=96067599; PubMed=7487939;  
 RA Boehm S.K., Gum J.R., Jr., Erickson R.H., Hicks J.W., Kim Y.S.;  
 RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a TATA-less GC-rich sequence characteristic of a housekeeping gene promoter.";  
 RL Biochem J. 312:835-843(1995).  
 RN [9]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Kidney;  
 RX MEDLINE=93210468; PubMed=8096237;  
 RA Morrison M.E., Vijayaradhni S., Engelstein D., Albino A.P.,  
 RA Houghton A.N.;  
 RT "A marker for neoplastic progression of human melanocytes is a cell surface ectopeptidase.";  
 RL J. Exp. Med. 177:1135-1143(1993).

CC - FUNCTION: Removes N-terminal dipeptides sequentially from  
 CC polypeptides having unsubstituted N-termini provided that the  
 CC penultimate residue is proline. Plays a role in T cell activation.  
 CC - CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-  
 CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided  
 CC Xcc is neither Pro nor hydroxyproline.  
 CC - SUBUNIT: Homodimer or heterodimer with Seprase (FAP).  
 CC - SUBCELLULAR LOCATION: Type II membrane protein. Also exists in  
 CC a soluble form.  
 CC - PTM: The soluble form (SDPP) derives from the membrane form (MDPP)  
 CC by proteolytic processing.  
 CC - SIMILARITY: Belongs to peptidase family S9B.  
 CC - DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: U13735; AAB60646.1; -;  
 CC EMBL: U13710; AAB60646.1; JOINED.  
 CC EMBL: U13711; AAB60646.1; JOINED.  
 CC EMBL: U13712; AAB60646.1; JOINED.  
 CC EMBL: U13713; AAB60646.1; JOINED.  
 CC EMBL: U13714; AAB60646.1; JOINED.  
 CC EMBL: U13715; AAB60646.1; JOINED.  
 CC EMBL: U13716; AAB60646.1; JOINED.  
 CC EMBL: U13717; AAB60646.1; JOINED.  
 CC EMBL: U13718; AAB60646.1; JOINED.  
 CC EMBL: U13719; AAB60646.1; JOINED.  
 CC EMBL: U13720; AAB60646.1; JOINED.  
 CC EMBL: U13721; AAB60646.1; JOINED.  
 CC EMBL: U13722; AAB60646.1; JOINED.  
 CC EMBL: U13723; AAB60646.1; JOINED.  
 CC EMBL: U13724; AAB60646.1; JOINED.  
 CC EMBL: U13725; AAB60646.1; JOINED.  
 CC EMBL: U13726; AAB60646.1; JOINED.  
 CC EMBL: U13727; AAB60646.1; JOINED.  
 CC EMBL: U13728; AAB60646.1; JOINED.  
 CC EMBL: U13729; AAB60646.1; JOINED.  
 CC EMBL: U13730; AAB60646.1; JOINED.  
 CC EMBL: U13731; AAB60646.1; JOINED.  
 CC EMBL: U13732; AAB60646.1; JOINED.  
 CC EMBL: U13733; AAB60646.1; JOINED.  
 CC EMBL: U13734; AAB60646.1; JOINED.  
 CC EMBL: M74777; AAA51943.1; -;  
 CC EMBL: BC013329; AAH13329.1; -;  
 CC EMBL: M80536; AAA52308.1; -;  
 CC EMBL: X60708; CAA43118.1; -;  
 CC EMBL: S79876; AAB55614.1; -;  
 CC PIR: S24313; CDH026.  
 CC PDB: INIM; 07-JAN-03.  
 CC MEROPS: S09.003; -;  
 CC Gene: HGNC:3009; DPP4.  
 CC MIM: 102720; -;  
 CC GO: GO:0004274; F:dipeptidyl-peptidase IV activity; TAS.  
 CC GO: GO:0005209; P:plasma protein; TAS.  
 CC GO: GO:0006955; P:immune response; TAS.  
 CC InterPro: IPR002469; DPP4V N term.  
 CC InterPro: IPR001375; Peptidase S9.  
 CC InterPro: IPR002471; ProL endopep. ser.  
 CC InterPro: IPR000379; Ser\_eactr\_site.  
 CC Pfam: PF00930; DPP4V\_N\_term; 1.  
 CC Pfam: PF00326; Peptidase S9; 1.  
 CC PROSITE: PS00708; PRO\_ENDOPEP\_SER; 1.  
 CC Hydrolase, Aminopeptidase; Dipeptidase; Serine protease;  
 CC Transmembrane; Glycoprotein; Signal-anchor; 3d-structure;  
 CC Dipeptidyl PEPTIDASE IV MEMBRANE FORM  
 CC CHAIN 1 766 (MDPP).

FT CHAIN 0 39 766 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM  
 FT DOMAIN 1 6 (SDPP).  
 FT TRANSMEM 7 28 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 29 766 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT ACT SITE 630 630 (POTENTIAL).  
 FT ACT SITE 708 708 EXTRACELLULAR (POTENTIAL).  
 FT ACT SITE 740 740 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 6 6 K -> R (IN REF. 6).  
 FT CONFLICT 7 7 V -> I (IN REF. 1).  
 FT CONFLICT 437 437 S -> I (IN REF. 1).  
 FT CONFLICT 557 557 T -> I (IN REF. 7).  
 FT CONFLICT 663 663 D -> E (IN REF. 7).

Query Match 19.9%; Score 333.5; DB 1; Length 766;  
 Best Local Similarity 27.7%; Pred. No. 8,3e-22;  
 Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;

QY 5 KSPLEHLLVYVYVNGEYRLTDGYSHSCIS-----QHCDFFISKYSNOKNP-- 56  
 DB 423 KGMFGGRNLVYK-----QLSD--YTKVLTCLSELNPERQYVSFSKAKYQLR 471  
 QY 57 C-----SLKLSPEDDPCKTEFEATILDSAGPLPDTPPE---IFSESTTGFLYG 109  
 DB 472 CGSGPLTYLLSHSVNDKGLRVLED--NSALDKM--LQNVQMPKSLDFTILNTEKF-WYQ 527  
 QY 110 MLYKPHDLQPKKYVFLVFIYGGP----- 133  
 DB 528 MLTPRH-PDKSKKYRLLDLNVAGPCQKADTYFRLLMATTYLAETMIIVASFGRGSGYQ 586  
 QY 134 -----OQOIEIDQVSGLYASRYFDIDRAGHWSYGYGLMALMQRSD 182  
 DB 587 GPKIMHAINRRLGTFEVEDQIEAARQF-SKMGVVDNRKRIALWMSYGVYTVSWLSCGSG 645  
 QY 183 IRRVALAGAPVTLMTITDGYTERYWG--HPDNEGGYLVGYAMAEKRPSPENRLLLL 240  
 DB 646 VPKCGIAPVAVSRWEYDVTERTYVGLTPEDNLDHYNSTVMSRAENF--KQVEYLLI 703  
 QY 241 HGFLENVHFATSTILSFLVAGKPYDLQIYQDERSIRVPESGHEYLHLHYLOE 298  
 DB 704 HGTADNVHFFQSAQISKALVDGVDFQAMWYTTDEDHGIASSTAHOHIYTHMSHTKQ 761

RESULT 7  
 ST13 YEAST STANDARD; PRT; 931 AA.  
 AC P33854; ST13 YEAST  
 DT 01-FEB-1994 (rel. 28, Created)  
 DT 01-FEB-1994 (rel. 28, Last sequence update)  
 DT 28-FEB-2003 (rel. 41, Last annotation update)  
 DE Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DBAP A) (YSCIV).  
 GN STE13 OR YC11 OR YOR219C OR YOR50-9.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OC NCB1\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9506382; PubMed=7975897;  
 RA Anna-Atticola S.S., Herskowitz I.;  
 RT "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl  
 RT aminopeptidase.";  
 RT Yeast 10:801-810(1994).

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB320;
RA Planagan C.A., Thorne J.;
RT "STB3";
RL (in) Getting M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL Press, Oxford (1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RA MEDLINE=96437977; PubMed=8840505;
RX Galleson F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae.";
RL Yeast 12:877-885(1996).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC ALPHA-FACTOR PRECURSOR.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: Belongs to peptidase family 93B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L21944; AAA35119.1; -
DR EMBL: U08230; AAA17897.1; -
DR EMBL: X92441; CAA63182.1; -
DR EMBL: X75127; CAA99437.1; -
DR PIR: A49737; A49737; -
DR MEROPS: S09.005; -
DR SCD; S0005745; STB3.
DR GO; GO:0005802; C:Golgi trans-face; IDA.
DR GO; GO:0004177; F:aminopeptidase activity; IDA.
DR GO; GO:0007323; P:peptide pheromone maturation; IDA.
DR InterPro: IPR002469; DPPIV N term.
DR InterPro: IPR001375; peptidase_S9.
DR InterPro: IPR002471; prol_endopep_ser.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR PROSITE; PS00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE NEG.
KW Hydrolyase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT DOMAIN 1 119
FT TRANSMEM 120 140
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 141 931
FT ACT_SITE 785 785 LUMENAL (POTENTIAL).
FT ACT_SITE 863 863 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 896 896 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 931 AA; 107200 MW; 81AF70094093C023 CRC64;

Query Match 18.7% Score 314; DB 1; Length 931;
Best Local Similarity 25.6%; Pred. No. 5; 9e-20;
Matches 100; Conservative 54; Mismatches 133; Indels 104; Gaps 12;

QY 1 FEETGKSPLEHLLVVSIV-----NPGEVRLTDRGYSHSCCISQHCDFRISK 48
DB 562 FTANIEIVMSQHLYSLTSTQNTQTFQSLQNP-----SDKYDFYDFELSSSRVATISK 615
QY 49 YSNQKN-----HCVALYKLSPEDDPTCKTKFMAITLDSACPLPDYT 92
DB 616 KLGDPPTIKVAGPLTRVLNVAEIHDDSLDQTKDE-----KFKX-----KINNYD 660
QY 93 PPEIFSESTT-----GFTLYGLMLKPHDLQPGKKYPTVLFYIGGQGG----- 136

```

```

DB 661 LP-ITSYKTVLDDGVEINYLIEIKPANLNKXKYPILVNIYGGSGQFTTKSLAFEQ 719
QY 137 -----LEIDDOVEG-----LOYLASRYDFIDRV 161
DB 720 VWSGLDVIYVQIETPRGCGWSFRSMAREKLGWEPRODTTEVTKKFIQNSCHIDSKI 779
QY 162 GHWMSYGYLSMAL-MQRSDFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNGCYL 220
DB 780 AIWMSYGFSTSLKTVELDNGDTFKYAMAVAPVTNNLTLYDSVYTERYMQPSEHNEGFE 839
QY 221 GSVAQAQKPEPPNNLLHGFLENVFAHLSILFLVRAG-KPYDQIYQENHSI 279
DB 840 VSTIQNFKSPES-LKRLFIHGTGFDNVAHQNTFRVLVDQNLGLTNVDMHIFPDSOHSI 898
QY 280 RVPESEFHEVLLHLYLOENLGSRIALAKYI 310
DB 899 RYHNAQRIYFQKLYWMLRDAFAERFNTFVL 929

RESULT 8
ID DAP2 YEAST STANDARD; PRT; 818 AA.
AC P18963;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPAF B) (YSCV).
GN DAP2 OR YHR028C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89174971; PubMed=2647766;
RA Roberts C.J., Fehlig G., Rothman J.H., Stevens T.H.;
RT "Structure, biosynthesis, and localization of dipeptidyl
RT aminopeptidase B, an integral membrane glycoprotein of the yeast
RT vacuole.";
RL J. Cell Biol. 108:1363-1373(1989).

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Giesel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Lareelle P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Ritkin L., Riles L., St Peter H., Tevaskis E., Vaughan K.,
RA Vagnati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:207-2082(1994).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: Belongs to peptidase family 93B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15484; CAA3512.1; -
DR EMBL: U10399; AAB68879.1; -
DR PIR: S46780; A30107.
DR MEROPS: S09.006; -
DR SCD; S0001070; DAP2.
DR GO; GO:0000329; C:vacuolar membrane (sensu Fungi); IDA.
DR InterPro: IPR002469; DPPIV_N_term.

```



```

DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; Prol_endopep_ser.
DR InterPro: IPR000379; Ser_gstrs_site.
DR Pfam: PF00330; Depr_N_term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
DR Hydrolase: Amino-peptidase; Dipeptidase; Serine protease;
KM Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 30 45 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 46 818 LOMENAL (POTENTIAL).
FT ACT_SITE 679 756 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 756 789 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 789 789 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 43 63 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 139 139 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 738 738 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 83 83 Q -> H (IN REF. 1).
FT CONFLICT 125 125 S -> N (IN REF. 1).
FT CONFLICT 125 188 FEELGNE -> LRRLT (IN REF. 1).
FT CONFLICT 182 200 D -> N (IN REF. 1).
FT CONFLICT 200 200 TSINWNESS -> DKRGRKEKF (IN REF. 1).
FT CONFLICT 366 375 AKRAFGQFVK -> QSVLSMGNLNLTLTYSSHRDIHKT
CONFLICT 808 818 FSYLHTMYI (IN REF. 1).
SQ SEQUENCE 818 AA; 93404 MW; 318F45045375BD3 CRC64;

Query Match 18.5%; Score 310; DB 1; Length 818;
Best Local Similarity 26.9%; Pred. No. 1,1e-19;
Matches 98; Conservative 38; Mismatches 122; Indels 106; Gaps 13;

QY 1 FEGRDSPLEHLLVYVSVNPGSEVRLTDRGYS--HSCCISQHCDFISKYSNOKNF-- 56
DB 451 FISTKKSSTERYVYIDLSRPNBIEVTDSQGVYVDFSSGRRRGLLTGKPKVYK 510
QY 57 -----C-----VSLYKSSPPDDPCTCKEWTATLDSAGLPYTPPEIFS 98
DB 511 IVDFFSRKAECDCGNVLGKSLYHLEKNE-----VLTQI--LEDYAVPRKSF 555
QY 99 PESTTGFTLYG--MLYKPHDLP-----GKKYPTVLFIYGQP----- 134
DB 556 RELNLGKDFEKKDILVNSYELLPNDFELTSHQYVFFFAVGGSPNSQVVKTFSSGENEV 615
QY 135 -----GQLEIDDQVEGLQYLASRYDFIDLRVG 162
DB 616 VASQLNATVYVVDGRGTGFKGQDFRSLVDRLDGYARQDIAASLYGS-LTFVDPQKIS 674
QY 163 IHGMSYGYGLSMALMQRSD-IFRVAIAGAPVTLFIYDTGYTERYMGHPDNEQGYLYG 221
DB 675 LFGMSYGYGLTLTKLEKDGSRHFKYGMVAPTDMFYSYTERYMHPTQENFDGYVBS 734
QY 222 SV-----AMGAKEKFPSEPNLLLLHGFLENNVFAHHSILSLFLVARG-KPYLQIYPOQR 276
DB 735 SVHNTAL-----AQAARFLMHGTDGNVHFQNSLKFLDLLDLNGENVYDVHVPDSD 788
QY 277 HSIR 280
DB 789 HSIR 792

```

```

RESULT 9
SEPR_MOUSE
ID SEPR_MOUSE STANDARD; PRT; 761 AA.
AC P97321;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral

```

```

DE membrane serine protease).
GN FAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_Taxid:10090;
RX MEDLINE=22388257; PubMed=9139873;
RA Niedermeyer J., Scanlan M.J., Garin-Chesa P., Daiber C., Fiebig H.H.,
RA Old L.J., Rettig W.J., Schnapp A.;
RT "Mouse fibroblast activation protein: molecular cloning, alternative
RT splicing and expression in the reactive stroma of epithelial
RT cancer.";
RL Int. J. Cancer 71:383-389 (1997).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=breast;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M., Usdin I.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.J., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak G.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=P97321-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P97321-2; Sequence=VSP_005368;
CC Name=3;
CC IsoId=P97321-3; Sequence=VSP_005369;
CC -1- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
CC embryos from day 7-19 and in new-born mice (P1).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: Y10007; CAA71116.1; -.
CC EMBL: BC019190; AAH19190.1; -.
CC DR MEROPS: S09.007; -.
CC MGD: MGI:109608; Fap.

```

```

DR InterPro: IPR002469; DPPIV_N term.
DR InterPro: IPR001375; Peptidase S9.
DR InterPro: IPR002471; Prol endopep ser.
DR InterPro: IPR000379; Ser_ester_site.
DR Pfam: PF00930; DPPIV_N_term; 1.
DR Pfam: PF00326; Peptidase S9; 1.
DR ProSite: PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 4
FT TRANSMEM 5 25
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 26 761
FT EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 31 35 Missing (in isoform 2).
FT VARSPPLIC /FTID=VSP_005368.
FT VARSPPLIC Missing (in isoform 3).
FT VARSPPLIC /FTID=VSP_005369.
FT CONFLICT 737 737 S -> L (IN REF. 2).
FT SEQUENCE 761 AA; 87944 MW; 9174CA3AEDA213B25 CRC64;

Query Match 18.4%; Score 309.5; DB 1; Length 761;
Best Local Similarity 25.6%; Pred. No. 1.1e-19;
Matches 91; Conservative 59; Mismatches 133; Indels 73; Gaps 10;

OY 1 FEGTDSPLFHLHYVSYVNPGEVTLTDRGYSHCSCISQH-----CDFEIKSYNSQKNP 55
DB 415 FEG---YFGRNITRISIGNSPPSK-----CVCCHLRKRCQYTTSPSKAKY 461
OY 56 HCVSLY---KLSPEDPTCKTEKFWATILDSAGLPDYPPET--FSFESTGTFTLYGM 110
DB 462 VALVCYGPGLPISTLHDSRTDOEIQVLEENKELENSLRNIQPKVYIKKLKDGGLTFWYK 521
OY 111 LYKPHLDQKQKPYVLYFYGGPQO--IETDOVESLQTLAR----- 152
DB 522 MLPPQDFRSKRYPLLIQVYGGPQSVKSVFAVNMVITTLAKREGIVTALVDGRGTAFOG 581
OY 153 -----YD-----FIDLDRVGIGHSGSYGLSLMLMGRSDIF 184
DB 582 DKFLAVYRKLGVYEVEDQLTAVRKTIEMGFIDEERIALWMSYGYVSLAASGTGLF 641
OY 185 RVALAGAPVTLMIFYDTGYTERYMGHP--DQNEQGYLGSVAMQAEKPPSEPNRLLLHNG 242
DB 642 KCGIYAVAPVSWMEYVASYISERFMGLPTKODNLEHKNSTVMAAEYFRNV--YLLING 699
OY 243 FLDENVHFMHTSILSLFVIRACKPYDLQIYPOERHSIRVPEGSEHMLLHYLOE 298
DB 700 TADNVHFQNSAQIALKALVNAQVDFQAMWYSQDNHISGSRQNHLYTHMTHFLKQ 755

RESULT 10
SEPR_HUMAN STANDARD; PRT; 760 AA.
AC Q12884; O00199; Q93998; Q90UD4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE membrane serine protease) (170-kDa melanoma membrane-bound
DE gelatinase).
GN FAP.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fibroblast;
RX MEDLINE=94261645; PubMed=7911242;
RA Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncasí M.P.,
RA Healey J.H., Old L.U., Rettig W.U.;
RT "Molecular cloning of fibroblast activation protein alpha, a member of
RT the serine protease family selectively expressed in stromal
RT fibroblasts of epithelial cancers.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Melanoma;
RX MEDLINE=97218181; PubMed=9247085;
RA Goldstein L.A., Gherzi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y.,
RA Plegasate D., Chen W.-T.;
RT "Molecular cloning of seprase: a serine integral membrane protease
RT from human melanoma.";
RL Biochim. Biophys. Acta 1361:11-19(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
RP 511-518.
RC TISSUE=Melanoma;
RX MEDLINE=97218181; PubMed=9065413;
RA Pineiro-Sanchez M.L., Goldstein L.A., Dodd J., Howard L., Yeh Y.,
RA Chen W.-T.;
RT "Identification of the 170-kDa melanoma membrane-bound gelatinase
RT (seprase) as a serine integral membrane protease.";
RL J. Biol. Chem. 272:7595-7601(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Melanoma;
RX MEDLINE=20112818; PubMed=10644713;
RA Goldstein L.A., Chen W.-T.;
RT "Identification of an alternatively spliced seprase mRNA that encodes
RT a novel intracellular isoform.";
RL J. Biol. Chem. 275:2554-2559(2000).
RN [5]
RP SEQUENCE OF 192-208; 220-240 AND 510-521.
RX MEDLINE=94327249; PubMed=7519584;
RA Rettig W.U., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
RA Garin-Chesa P., Healey J.H., Old L.U.;
RT "Fibroblast activation protein: purification, epitope mapping and
RT induction by growth factors.";
RL Int. J. Cancer 58:385-392(1994).
CC -!- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and may contribute to invasiveness in malignant
CC cancers.
CC -!- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -!- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
CC surface lamellipodia, invadopodia and on shed vesicles.
CC -!- ALTERNATIVE PRODUCTS:
CC Name=1; Synonyms=L;
CC IsoId=Q12884-1; Sequence=Displayed;
CC Note=Major isoform;
CC Name=2; Synonyms=S, truncated;
CC IsoId=Q12884-2; Sequence=VSP_005367;
CC -!- TISSUE SPECIFICITY: Fibroblast-specific.
CC -!- INDUCTION: In fibroblasts at times and sites of tissue remodeling
CC during development, tissue repair, and carcinogenesis.
CC -!- PTM: N-glycosylated.
CC -!- PTM: The N-terminus may be blocked.
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; U09278; AAB49652.1; -  
 CC EMBL; U76833; AAC51668.1; -  
 CC EMBL; AF007822; AAF21600.1; -  
 CC MEROPS; S09.007; -  
 CC Genew; HGNC:3590; FAP.  
 DR MIM; 600403; -  
 DR InterPro; IPR002469; DDPV\_N term.  
 DR InterPro; IPR001375; Peptidase\_S9.  
 DR InterPro; IPR002471; ProI endopep ser.  
 DR InterPro; IPR000379; Ser esters site.  
 DR Pfam; PF00930; DDPV\_N term; 1.  
 DR Pfam; PF00326; Peptidase\_S9; 1.  
 DR ProSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
 KW Hydrolyase; Protease; Serine protease; Transmembrane; Signal-anchor;  
 KW Glycoprotein; Alternative splicing.  
 FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 26 760 EXTRACELLULAR (POTENTIAL).  
 FT ACT SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1 521 Missing (in isoform 2).  
 FT FTID=VSP\_005367.  
 FT P -> A (IN REF. 2).  
 FT K -> T (IN REF. 2).  
 FT R -> T (IN REF. 2).  
 FT CONFLICT 207 207  
 FT CONFLICT 229 229  
 FT CONFLICT 354 354  
 SQ SEQUENCE 760 AA; 87820 MW; A0D34B4801BE07EA CRC64;

Query Match 18.0%; Score 302; DB 1; Length 760;

Best Local Similarity 25.5%; Pred. No. 5,3e-19; Matches 93; Conservative 55; Mismatches 139; Indels 78; Gaps 11;

1 FEETKDSPLEHLYVSVVPGVETRLTDRGYSHS-CCISQH-----CDFEIKSYSNOKN 54  
 401 FRTQDSLFSSNSPEFEYRGRNIIYISIGSYPPSKCVCHLXKRCQYVYASFDYAK 460  
 55 PHCVSLY-----KLSPEDDPTCKEFAWATILDSAGPLD-----YTPPEIFSESTIGF 105  
 461 YVALVCYGGPISTLHDGRTRDQ-----IKILEENKELEALKNIQLPKEIKLEVD 516  
 106 TLVGMLYKHDLQPGKKYPTVLFLTYGSP----- 133  
 517 TLVYKMLLPQGFDRSKKYPILLVYGGPQSCSVYFAVNMISYLASKEGMVALVDGRG 576  
 134 -----OGQIEIDQVEGLQYLASRYFDIDRVYGIHMSYGYLSMALM 178  
 577 TAFQDKLAVYVRKLGVEVEEQITAVRKFT-EMGFIDSKRLAIMGWSGGVYSSIALA 635  
 179 QRSDFRVAJAGAPVTLMIFDYTGTERVYGNP--DQNEGGYLVGSVAMAQKFPSEPNR 236  
 636 SGTGLFKCGIAVAPVSMWEYVASVTERFGLPTKDNLEHYKNSVTMAAEYFRVND-- 693  
 237 LLLHGFDENVHFAHTSILSLFLVRAAGKRYDQIYPOEHNSIRVPSG---EHYELHL 293  
 694 YLLIHGTADNVHFOHNSAQIAKALVNAQVDFQAMWYSDQNHGL-----SGLSHLYTHMT 749  
 294 HYLQE 298  
 750 HFLQK 754

RESULT 11  
 DPP6\_BOVIN

ID DPP6\_BOVIN STANDARD; PRT; 863 AA.

AC PA2659;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)  
 DE Dipeptidyl aminopeptidase 6 (Dipeptidyl peptidase IV like protein)  
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).  
 OS DPP6.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain;  
 RA MEDLINE=92108018; PubMed=1729689;  
 RA Mada K., Yokotani N., Hunter C., Doi K., Wentholt R.J., Shimasaki S.;  
 RT "Differential expression of two distinct forms of mRNA encoding  
 RT members of a dipeptidyl aminopeptidase family.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN  
 CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF  
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Dppx-L;  
 CC IsoId=P42659-1; Sequence=Displayed;  
 CC Name=Dppx-S;  
 CC IsoId=P42659-2; Sequence=VSP\_005364;  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L  
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN  
 CC BRAIN, KIDNEY, OVARY AND TESTIS.  
 CC -1- SIMILARITY: belongs to peptidase family S9B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M76428; AAC41622.1; -  
 CC EMBL; M76429; AAC41623.1; -  
 CC PIR; A41793; A41793.  
 CC MEROPS; S09.973; -

DR InterPro; IPR002469; DDPV\_N term.  
 DR InterPro; IPR001375; Peptidase\_S9.  
 DR InterPro; IPR000379; Ser esters site.  
 DR Pfam; PF00930; DDPV\_N term; 1.  
 DR Pfam; PF00326; Peptidase\_S9; 1.  
 KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.  
 FT DOMAIN 1 93  
 FT TRANSMEM 94 114  
 FT (POTENTIAL).

FT DOMAIN 115 863  
 FT CARBOHYD 171 171  
 FT CARBOHYD 402 402  
 FT CARBOHYD 469 469  
 FT CARBOHYD 533 533  
 FT CARBOHYD 564 564  
 FT CARBOHYD 811 811  
 FT VARSPLIC 1 79

FT FTID=VSP\_005364.  
 FT P -> A (IN REF. 2).  
 FT K -> T (IN REF. 2).  
 FT R -> T (IN REF. 2).  
 FT CONFLICT 207 207  
 FT CONFLICT 229 229  
 FT CONFLICT 354 354  
 SQ SEQUENCE 863 AA; 96556 MW; 23DBA7928B41A39D CRC64;

Query Match 15.6%; Score 262; DB 1; Length 863;

```

Best Local Similarity 25.0%; Pred. No. 2,3e-15;
Matches 89; Conservative 45; Mismatches 152; Indels 70; Gaps 9

OY      1 FEGTQDSPLEHLLVYVSYNPEEVR-----LDRGSHSCCISQCHDFISKYSNQK 54
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      500 FLSTEDLPRRROLVASTV--GSFNROCLSCDLVNCYFASFSFGADFFLLKCEGCV 557
OY      55 PCHVELYKLSSPEDDPTCKTKE-FWATILDSAGPLDTPPEIFSFESTGTGLGMLYK 113
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      558 P-TVSVHNTDKKKMFDELTEHNVOKAISDRQPKVEYRKLE-----TDYINLPQILK 610
OY      114 PHDLQPKKKYPTPLVIYGGP-----LTKK----- 133
DB      611 PATFTDTHAYPELLVADGTPGSGVAEKAQVWETVMVSSHGAVVVKCDRGSGFGRTLR 670
OY      134 -----OGQEIFDDVQESLOYLARVDFIDLDRIHGIMSGVGLSLMALMQBSD---- 182
DB      671 LHEVRRRLGSLSEKQOMEAVRWML-KEPTIDTRAVAGKQDQVGLSTYLLLPKADGGAP 729
OY      183 IFRVAIAGAPVLWLFIDTGYTERYMGHPDQEGYIGSVAMAQAEKPEEENRLLTHG 242
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      730 VFSGGSLSPITDFLTVASAFERILGLGDNRAVEAKVAHRVSL--EQQGLVYIHA 787
OY      243 FLDENVHFAMHSILSFLVRACKKPYDLOQYPOERHSIRVPESGEYELHLHYLOE 298
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      788 TADEKIHQHTAEILLTOLIKGRANVSLQIYPPESHVFSSALQHLHRSILGFVE 843

RESULT 12
DPP6_HUMAN
1D_DPP6_HUMAN STANDARD; PRT; 865 AA.
AC P42658;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
DE DPP6.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid=9606;
RX MEDLINE=93372805; PubMed=8103397;
RA Yokotani N., Doi K., Wenthold R.J., Wada K.;
RT "Non-conservation of a catalytic residue in a dipeptidyl
aminopeptidase IV-related protein encoded by a gene on human
chromosome 7.";
RL Hum. Mol. Genet. 2:1037-1039(1993).
CC - FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC - SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC - ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=DPPX-L;
CC IsoId=P42658-1; Sequence=Displayed;
CC Name=DPPX-S;
CC IsoId=P42658-2; Sequence=VSP_005365;
CC - TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC - SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
```

Query Match	14.8%	Score 248	DB 1	Length 865
Beat Local Similarity	23.0%	Pred. No. 4.1e-14		
Matches 84	Conservative 53	Mismatches 139	Indels 90	Gaps 10
QY 1	FEETKDSPLEHLLVSVYNDGEVRLTDRGYSH--SCCISQHCDFEISKYSNOKMPHC	57		
DB 502	FLSTEDLPRRLQVLSAN-----TEGNFNQCISCLDVLNCTYFSAFSHSM	548		
QY 58	VSIVKLSR-----EDDPTCKTEF-----WATILDSAGLPLDPTPEISFESTT	103		
DB 549	FPLKCEGPGVPMVTNHTTIDKKMFPLETNEHVKKAIINDQMPKVEYRDI	602		
QY 104	GFTLYGMLYKHEHDQPKKYPTVLFYIGGPGQ-----	136		
DB 603	DYNLPMLIKRATGTTDTTHRYPLLVDGTPSSQSAVEKEFEVSMETWVSSHGAVVKKDCG	662		
QY 137	-----IEDDQVEGLQYLAISRYDFIDLDRVGIHGSYGYLSMA	176		
DB 663	RSGSGFOGTKLHEVRRRLGLLEEDQOMEAVRTM--KEQYIDRTVAVAGKQYGYLSTYI	721		
QY 177	LMORS-----IFRVAIAGAPVTLTIFPDQGTETMYMGHPDQEGGYLIGVAMAEKRP	232		
DB 722	LPAKGENOGOTFTGCSALSPITDCKLYASAFSEYLLGHGIDNAYETTKVAHRVSA--	779		
QY 233	EPNNLLLLGLFIDENVAFAHTSILSEFLVAKGPEYDQIQPOEHSIRVPSGSEHYELHT	292		
DB 780	EEQGFLLIHPADEKIHFOHTAELITOLIRGKANYSIQIYDDESHYFTSSLKQHLVRSI	839		
QY 293	LHYLQE 298			
DB 840	INFEVE 845			





Db	777	QPLIHATADEKIHFOHTAELITOLIKGKANYSLQIYPDESHYFHSVALKOHLYRSIIGF	836
Qy	296	LQE	298
Db	837	FVE	839

Search completed: October 15, 2003, 17:11:28  
Job time : 6.9182 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 ; Search time 19.8265 Seconds

(without alignments)  
4034.822 Million cell updates/sec

Title: US-10-070-464-3  
Perfect score: 1680  
Sequence: 1 FEGTKDSPLEHLYVSVYN.....HLHYLQENLGRNALAKVI 310

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_proteus:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvivirus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	1680	100.0	310	4 Q9HBM4	Q9HBM4 homo sapien
2	1645.5	97.9	882	4 Q9HBM5	Q9HBM5 homo sapien
3	1645.5	97.9	882	4 Q8HGM7	Q8HGM7 homo sapien
4	1605.5	95.6	892	11 Q9D4G6	Q9D4G6 mus musculu
5	1407	83.8	831	4 Q8NEM5	Q8NEM5 mus musculu
6	1401	83.4	632	4 Q9GJX1	Q9GJX1 homo sapien
7	1377.5	82.0	465	4 Q9HBM3	Q9HBM3 homo sapien
8	1094.5	65.1	862	11 Q8HWT9	Q8HWT9 mus musculu
9	1093.5	65.1	439	4 Q9BVR3	Q9BVR3 homo sapien
10	1093.5	65.1	862	11 Q8BVG4	Q8BVG4 mus musculu
11	1093.5	65.1	863	4 Q8WMD8	Q8WMD8 homo sapien
12	1086.5	64.7	360	4 Q9HBM2	Q9HBM2 homo sapien
13	1032.5	57.6	508	4 Q7S273	Q7S273 homo sapien
14	968.5	57.6	628	4 Q8N3F5	Q8N3F5 homo sapien
15	938.5	55.9	312	4 Q9GNT8	Q9GNT8 homo sapien
16	708	42.1	1042	5 Q9VC20	Q9VC20 drosophila

17	708	42.1	1102	5 Q9VC19	Q9VC19 drosophila
18	708	42.1	1053	5 Q8IH07	Q8IH07 drosophila
19	659	39.2	167	4 Q8N2J7	Q8N2J7 homo sapien
20	502	29.9	927	5 Q965K3	Q965K3 caenorhabdi
21	502	29.9	931	5 Q44987	Q44987 caenorhabdi
22	484.5	28.8	738	16 Q9A6E0	Q9A6E0 caulobacter
23	442	26.3	552	10 Q8GJY7	Q8GJY7 arabidopsis
24	442	26.3	746	10 Q9FNF6	Q9FNF6 arabidopsis
25	441.5	26.3	757	16 Q8FPD7	Q8FPD7 xanthomonas
26	436.5	26.0	753	16 Q8EAB7	Q8EAB7 shewanella
27	433.5	25.8	751	16 Q8P3V8	Q8P3V8 xanthomonas
28	398.5	23.7	711	2 P95782	P95782 xanthomonas
29	379	22.6	711	2 Q47900	Q47900 flavobacter
30	376	22.4	723	2 Q31048	Q31048 porphyromon
31	376	22.4	723	2 Q66223	Q66223 porphyromon
32	374.5	22.3	730	2 Q93JY4	Q93JY4 prevotella
33	343.5	20.4	901	3 Q96VT7	Q96VT7 aspergillus
34	341	20.3	748	13 P70092	P70092 xenopus lae
35	334	19.9	237	2 Q87543	Q87543 capnocytoph
36	328	19.5	802	5 Q9VTH1	Q9VTH1 drosophila
37	327	19.5	745	5 Q9VNM2	Q9VNM2 drosophila
38	327	19.5	771	3 Q42812	Q42812 aspergillus
39	315.5	18.8	765	3 Q14425	Q14425 aspergillus
40	314.5	18.7	761	11 Q8R492	Q8R492 rattus norv
41	305.5	18.2	935	5 Q9VMB4	Q9VMB4 drosophila
42	304.5	18.1	827	16 Q8PPU4	Q8PPU4 xanthomonas
43	303.5	18.1	799	5 Q18119	Q18119 caenorhabdi
44	293	17.4	707	16 Q9F348	Q9F348 streptomyce
45	284.5	16.9	829	5 Q18253	Q18253 caenorhabdi

## ALIGNMENTS

RESULT 1	Q9HBM4	PRELIMINARY;	PRT;	310 AA.
ID	Q9HBM4			
DT	01-MAR-2001 (Tremblrel. 16, Created)			
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	Dipeptidyl peptidase 8 (Fragment).			
GN	PPP8.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Butiria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=20467194; PubMed=11012666;			
RA	Abdott C.A., Yu D.W.T., Woollett E., Sutherland G.R., McCaughan G.W.,			
RA	Goirell M.D.;			
RT	"Cloning, expression and chromosomal localization of a novel human			
RT	dipeptidyl peptidase (DPP) IV homolog, DPP8.";			
RL	Eur J Biochem. 267:6140-6150(2000).			
DR	EMBL; AF221635; AAC29767.1; ..			
DR	MEKOPS; S09.018; ..			
DR	InterPro; IPR001375; Peptidase_S9.			
FT	Pfam; PF00326; Peptidase_S9; 1.			
FT	NON_TER			
SO	SEQUENCE			
Query Match	310 AA; 35396 MW; 8887C34026D9C7AC CRC64;			
Best Local Similarity	100.0%; Pred. No. 9.7e-145;			
Matches 310; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 FEGTKDSPLEHLYVSVYNPEGYRLTRDGRSHSCCTSOHCDPFIKSNQKPRCVSL 60			
DB	1 FEGTKDSPLEHLYVSVYNPEGYRLTRDGRSHSCCTSOHCDPFIKSNQKPRCVSL 60			
QY	61 YKLSSPEDPTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLKKPHDLOG 120			



```

Db      61 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPELFSFESTGTLLYGMLYKPHDLOPG 120
      121 KKPYTULFIYGGPOGQOIEIDDOVEGLOYLASRYDFIDLDVGIGHGMSYGYLSMALMOR 180
      121 KKPYTULFIYGGPOGQOIEIDDOVEGLOYLASRYDFIDLDVGIGHGMSYGYLSMALMOR 180
Qy      181 SDIFRVAIAGAPYTLWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLL 240
      181 SDIFRVAIAGAPYTLWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLL 240
Db      241 HGLDENVHFAHNSILSLFLVRACKPFDLOIYPOERSINVPESGEHYELHLHYLOENL 300
      241 HGLDENVHFAHNSILSLFLVRACKPFDLOIYPOERSINVPESGEHYELHLHYLOENL 300
Qy      301 GSRIALAKVI 310
      301 GSRIALAKVI 310
      301 GSRIALAKVI 310

```

## RESULT 2

```

AC      09HBM5; PRELIMINARY; PRT; 882 AA.
ID      09HBM5;
AC      01-MAR-2001 (T-EMBLrel. 16, Created)
DT      01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT      01-MAR-2001 (T-EMBLrel. 23, Last annotation update)
DE      Dipeptidyl peptidase 8.
GN      DPP8.
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RX      MEDLINE=20467194; Pubmed=11012666;
RA      Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA      Gorell M.D.;
RT      "Cloning, expression and chromosomal localization of a novel human
RT      dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL      Eur. J. Biochem. 267:6140-6150(2000).
EMBL: AF221634; AAC29766.1; -.
DR      MEROPS, S09_018; -.
DR      InterPro, IPR002469; DPP1V_N_term.
DR      InterPro, IPR001375; Peptidase_S9.
DR      InterPro, IPR000379; Ser_estra_site.
DR      Pfam, PF00930; DPP1V_N_term; 1.
DR      Pfam, PF00326; Peptidase_S9; 1.
SQ      SEQUENCE 882 AA, 101421 MW, AD801C302DB4652B CRC64;

```

```

Query Match      97.9%; Score 1645.5; DB 4; Length 882;
Best Local Similarity 86.4%; Pred. No. 5.5e-141;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```

```

Qy      1 FEETKOSPLEHHLVYVSYVNPGEVTRLTDGYSHSCISQHCDFPISKYSNQNPHCVSL 60
      524 FEETKOSPLEHHLVYVSYVNPGEVTRLTDGYSHSCISQHCDFPISKYSNQNPHCVSL 583.
Qy      61 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPELFSFESTGTLLYGMLYKPHDLOPG 120
      584 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPELFSFESTGTLLYGMLYKPHDLOPG 643
Qy      121 KKPYTULFIYGGPO----- 134
      644 KKPYTULFIYGGPOVOLLNRRFKGVKFRNLNTLASLGYVVVVIDNRGSCHRGLKFEAGFK 703
Qy      135 ----GQIEIDDOVEGLOYLASRYDFIDLDVGIGHGMSYGYLSMALMORSDFRVAIAGA 191
      704 YKMGQIEIDDOVEGLOYLASRYDFIDLDVGIGHGMSYGYLSMALMORSDFRVAIAGA 763
Qy      192 PVTLMIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLFDENVHFA 251
      764 PVTLMIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLFDENVHFA 823

```

```

Qy      252 HTSILSLFLVRACKPFDLOIYPOERSINVPESGEHYELHLHYLOENLGSRIALAKVI 310
      824 HTSILSLFLVRACKPFDLOIYPOERSINVPESGEHYELHLHYLOENLGSRIALAKVI 882

```

## RESULT 3

```

ID      081WG7 PRELIMINARY; PRT; 882 AA.
AC      081WG7;
DT      01-MAR-2003 (T-EMBLrel. 23, Created)
DT      01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE      Similar to dipeptidyl peptidase 8.
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RX      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Strausberg R.;
RL      Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL: BC040204; AAH40203.1;
SQ      SEQUENCE 882 AA, 101391 MW, 88C76AF5BCE707F9 CRC64;

```

```

Query Match      97.9%; Score 1645.5; DB 4; Length 882;
Best Local Similarity 86.4%; Pred. No. 5.5e-141;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```

```

Qy      1 FEETKOSPLEHHLVYVSYVNPGEVTRLTDGYSHSCISQHCDFPISKYSNQNPHCVSL 60
      524 FEETKOSPLEHHLVYVSYVNPGEVTRLTDGYSHSCISQHCDFPISKYSNQNPHCVSL 583
Qy      61 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPELFSFESTGTLLYGMLYKPHDLOPG 120
      584 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPELFSFESTGTLLYGMLYKPHDLOPG 643
Qy      121 KKPYTULFIYGGPO----- 134
      644 KKPYTULFIYGGPOVOLLNRRFKGVKFRNLNTLASLGYVVVVIDNRGSCHRGLKFEAGFK 703
Qy      135 ----GQIEIDDOVEGLOYLASRYDFIDLDVGIGHGMSYGYLSMALMORSDFRVAIAGA 191
      704 YKMGQIEIDDOVEGLOYLASRYDFIDLDVGIGHGMSYGYLSMALMORSDFRVAIAGA 763
Qy      192 PVTLMIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLFDENVHFA 251
      764 PVTLMIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLFDENVHFA 823
Qy      252 HTSILSLFLVRACKPFDLOIYPOERSINVPESGEHYELHLHYLOENLGSRIALAKVI 310
      824 HTSILSLFLVRACKPFDLOIYPOERSINVPESGEHYELHLHYLOENLGSRIALAKVI 882

```

## RESULT 4

```

ID      09D4G6 PRELIMINARY; PRT; 892 AA.
AC      09D4G6;
DT      01-JUN-2001 (T-EMBLrel. 17, Created)
DT      01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE      Adult male testis cDNA, RIKEN full-length enriched library,
DE      clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.
OS      Mus musculus (mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RX      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=Testis;
RA      Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA      Arai T., Bono H., Carninci P., Fukuda S., Fukushima Y., Furuno M.,
RA      Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,

```

RA Imotani K., Iehi Y., Itch M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Oka C., Saito K., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RA Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning."  
 RL Mech. Enzymol. 303:19-44(1999).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itch M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itch M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itch M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsui S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:1757-1771(2000).  
 DR EMBL; AK016546; BAB30295.2; -  
 SQ SEQUENCE 892 AA; 102284 MW; FDE0DBEDCA4CA346 CRC64;

Query Match 95.6%; Score 1605.5; DB 11; Length 892;  
 Best Local Similarity 84.1%; Pred. No. 2.5e-137;  
 Matches 302; Conservative 3; Mismatches 5; Indels 49; Gaps 1;

QY 1 FEGTKDSPLEHLLVVSIVNPGEVRLTDRGYSGCCISOCHDFISKSNOKNPHCVSL 60  
 DB 534 FEGTKDSPLEHLLVVSIVNPGEVRLTDRGYSGCCISOCHDFISKSNOKNPHCVSL 593  
 QY 61 YKLSPPEDPTCKTEKFEFATILDSAGPLDPTPEIFSFESTTGTLYGMLYKPHDLPG 120  
 DB 594 YKLSPPEDPTCKTEKFEFATILDSAGPLDPTPEIFSFESTTGTLYGMLYKPHDLPG 653  
 QY 121 KKYPTVLFYGGPO----- 134  
 DB 654 KKYPTVLFYGGPOVQLVNNRFGVKYFRLNTLASLGYVVVVVDNRGSGHGLKFECAFK 713

QY 135 ---GGIETDDQVEGIQYLASRYDFIDLVRVGHGSGYGLSLMALMQRSDIFRVAIAGA 191  
 DB 714 YKMGIEIDDDQVEGIQYLASRYDFIDLVRVGHGSGYGLSLMALMQRSDIFRVAIAGA 773  
 QY 192 PVTLMIFDYDTGTERYMGHPDNEGGYIGSVAMQAEKFPSPNNLLHGLDENVHFA 251  
 DB 774 PVTLMIFDYDTGTERYMGHPDNEGGYIGSVAMQAEKFPSPNNLLHGLDENVHFA 833  
 QY 252 HTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHLHYQENIGSRILAKVI 310  
 DB 834 HTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHLHYQENIGSRILAKVI 892

## RESULT 5

ID Q8NEM5 PRELIMINARY; PRT; 831 AA.  
 AC Q8NEM5;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Similar to dipeptidyl peptidase 8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 RN NCBI\_TaxID=9606;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Streueberg R.;  
 RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BC030688; AAH30688.1; -  
 SQ SEQUENCE 831 AA; 93527 MW; 0B2A13A2FE70CBE2 CRC64;

Query Match 83.8%; Score 1407; DB 4; Length 831;  
 Best Local Similarity 84.9%; Pred. No. 2.7e-119;  
 Matches 270; Conservative 11; Mismatches 19; Indels 18; Gaps 5;

QY 1 FEGTKDSPLEHLLVVSIVNPGEVRLTDRGYSGCCISOCHDFISKSNOKNPHCVSL 60  
 DB 524 FEGTKDSPLEHLLVVSIVNPGEVRLTDRGYSGCCISOCHDFISKSNOKNPHCVSL 583  
 QY 61 YKLSPPEDPTCKTEKFEFATILDSAGPLDPTPEIFSFESTTGTLYGMLYKPHDLPG 120  
 DB 584 YKLSPPEDPTCKTEKFEFATILDSAGPLDPTPEIFSFESTTGTLYGMLYKPHDLPG 643  
 QY 121 KKYPTVLFYGGPOQIETDDQVEGLQY----LAS-RYDFIDLVRGI--HGWSYGYL 172  
 DB 644 KKYPTVLFYGGPOQIETDDQVEGLQY----LAS-RYDFIDLVRGI--HGWSYGYL 702  
 QY 173 SLMALMQRSDIFRVAIAGAPVTLMIFYDTGTERYMGHPDNEGGYIGSVAMQAEKFP 232  
 DB 703 KYK-----MVAIAGAPVTLMIFYDTGTERYMGHPDNEGGYIGSVAMQAEKFP 753  
 QY 233 EPNRLLLHGLDENVHFAHTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHL 292  
 DB 754 EPNRLLLHGLDENVHFAHTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHL 813  
 QY 293 LHYQENIGSRILAKVI 310  
 DB 814 LHYQENIGSRILAKVI 831

## RESULT 6

ID Q96JX1 PRELIMINARY; PRT; 632 AA.  
 AC Q96JX1;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Hypothetical protein FJ14920.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Db 694 NMGQVEIEDQVEGLQVAAEKYGFIDLSRVAIHGWSYGFSLMGLIHKEQVFAIAGA 753  
 QY 192 PVTLMIFYDTGYTERWYGHNDQEGYLLGSVAMQAEKFPSEPRRLILLHGFIDENVHFA 251  
 Db 754 PVTVMAYDGYTERYMDVPENNQOGEASVALHVEKLPNEPRRLILLHGFIDENVHFF 813  
 QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENTL 300  
 Db 814 HTMFVLSQLIRAGKPYDLOIYPERHSIRCRESEGEHYELHLHYLOENTL 862

## RESULT 9

Q9BVR3 PRELIMINARY; PRT; 439 AA.  
 AC Q9BVR3;  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein (fragment).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 DB EMBL, BC00970; AA00970.1; -.  
 DR MEROPS; S09.019; -.  
 DR Genew; HGNC:18648; DPP9.  
 DR InterPro; IPR001375; Peptidase\_S9.  
 DR InterPro; IPR000379; Ser esterase.  
 DR Pfam; PF00326; Peptidase\_S9; 1.  
 KM Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 439 AA; 45926 MW; A18BA9E12092BAF CRC64;

Query Match 65.1%; Score 1093.5; DB 4; Length 439;  
 Best Local Similarity 57.3%; Pred. No. 4e-91;

Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 1 FEETKDSPLEHNLVYVSYNPGVETRLTDRGYSHSCISQHCDFIFISKYSNQNPHCVSL 60  
 Db 91 FQGTKDTPLEHNLVYVSYNPGVETRLTDRGYSHSCISQHCDFIFISKYSNQNPHCVSL 150  
 QY 61 YKLSSEDDPTCKTKERMAITLDSAGPLPYTPPEISFESTTGFTLYGMLYKPHDLOG 120  
 Db 151 YKLSGPDDEPLHQPFRFASMEANCPDPYVPEIFHFTRADVRLYGMITYKPHDLOG 210  
 QY 121 KKYPTVLFYIGFQ----- 134  
 Db 211 KKHPTVLFYIGFQVQLVNNSPFGIKYLRNTLASLGYAVVITDGRSCORGLHFEALK 270  
 QY 135 ---GQIEIDQVEGLQVLAARYDFIDLDRVGHGWSYGYLSMALMQRSDFRVAIAGA 191  
 Db 271 NMGQVEIEDQVEGLQVAAEKYGFIDLSRVAIHGWSYGFSLMGLIHKEQVFAIAGA 330  
 QY 192 PVTLMIFYDTGYTERWYGHNDQEGYLLGSVAMQAEKFPSEPRRLILLHGFIDENVHFA 251  
 Db 331 PVTVMAYDGYTERYMDVPENNQOGEASVALHVEKLPNEPRRLILLHGFIDENVHFF 390  
 QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENTL 300  
 Db 391 HTMFVLSQLIRAGKPYDLOIYPERHSIRCRESEGEHYELHLHYLOENTL 439

## RESULT 10

Q8BVG4 PRELIMINARY; PRT; 862 AA.  
 AC Q8BVG4;  
 DT 01-MAR-2003 (TEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Dipeptidyl peptidase 9 homolog.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Olfactory brain;  
 RX MEDLINE=22354683; PubMed=1246851;  
 RA The FANTOM Consortium;  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK078301; BAC37211.1; -.  
 SQ SEQUENCE 862 AA; 98049 MW; B4FD3476B6F50030 CRC64;

Query Match 65.1%; Score 1093.5; DB 11; Length 862;  
 Best Local Similarity 57.3%; Pred. No. 1e-90;  
 Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 1 FEETKDSPLEHNLVYVSYNPGVETRLTDRGYSHSCISQHCDFIFISKYSNQNPHCVSL 60  
 Db 514 FQGTKDTPLEHNLVYVSYNPGVETRLTDRGYSHSCISQHCDFIFISKYSNQNPHCVSL 573  
 QY 61 YKLSSEDDPTCKTKERMAITLDSAGPLPYTPPEISFESTTGFTLYGMLYKPHDLOG 120  
 Db 574 YKLSGPDDEPLHQPFRFASMEANCPDPYVPEIFHFTRADVRLYGMITYKPHDLOG 633  
 QY 121 KKYPTVLFYIGFQ----- 134  
 Db 634 KKHPTVLFYIGFQVQLVNNSPFGIKYLRNTLASLGYAVVITDGRSCORGLHFEALK 693  
 QY 135 ---GQIEIDQVEGLQVLAARYDFIDLDRVGHGWSYGYLSMALMQRSDFRVAIAGA 191  
 Db 694 NMGQVEIEDQVEGLQVAAEKYGFIDLSRVAIHGWSYGFSLMGLIHKEQVFAIAGA 753  
 QY 192 PVTLMIFYDTGYTERWYGHNDQEGYLLGSVAMQAEKFPSEPRRLILLHGFIDENVHFA 251  
 Db 754 PVTVMAYDGYTERYMDVPENNQOGEASVALHVEKLPNEPRRLILLHGFIDENVHFF 813  
 QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENTL 300  
 Db 814 HTMFVLSQLIRAGKPYDLOIYPERHSIRCRESEGEHYELHLHYLOENTL 862

## RESULT 11

Q8MXD8 PRELIMINARY; PRT; 863 AA.  
 AC Q8MXD8;  
 DT 01-MAR-2002 (TEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Dipeptidyl peptidase 9.  
 GN DPP9.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC Olgren C.; Wagmann N.;  
 RA "Identification and characterization of a novel member of the  
 RT dipeptidyl peptidase IV-related family."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 (2)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Strausberg R.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF452102; AAL47179.1; -.  
 DR EMBL; BC037948; AA037948.1; -.



```

OY 252 HTSILSLFLVAGKRPYD-----QIYQERHSIRVPSGEHYEHLHLHYLOENL 300
DB 451 HTNFLVSQLIRAGKRYQLQVALPVPSPQYPPNERHSIRCPESGEHYEYTLHLFLQEYL 508

RESULT 14
OBN3F5 PRELIMINARY; PRT; 628 AA.
AC OBN3F5;
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP762M2413.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Ansoorge W., Winkner U., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL8843376; CAD39039.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF00326; Peptidase_S9; 1.
DR Hypothetical protein.
KM NON TER
FT
SQ SEQUENCE 628 AA; 71368 MW; DC0B6A9440507C3B CRC64;

Query Match 57.6%; Score 968.5; DB 4; Length 628;
Best Local Similarity 55.1%; Pred. No. 1.6e-79;
Matches 177; Conservative 40; Mismatches 55; Indels 49; Gaps 1;

OY 1 PEGTKDSLEHLHLVYSVNPGEVRLTDRGYSHSCCISQCHDFISKYNQKNPHCVSL 60
DB 307 FQGRKDPLEHLHYVSYEAAGEIYRLTTPGSHSCSOSNFMDSVSHYSVSTPPCVH 366
OY 61 YKLSPPEDPTCKTEKFPATILDSAGPLPDYTPPEIFSESTGTLYGMLYKPHDLOPG 120
DB 367 YKLSGPDDDLHKOPRFWASMEASCPDYVPPPIFHFTRSDVRLYGMLYKPHALOPG 426
OY 121 KKYPYTLFLYGGP----- 134
DB 427 KKHPTVLFLVYGQVQLVNNSEFKIKYLRNLTLASLGAVVVIDGRSGCQGLRFEGLX 486
OY 135 ---GQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGGYLSLWLMQSRDIFRYAIGA 191
DB 487 NQMQQVEIEDQVEGLQYAEKRYGFDLSRVAIHGWSYGGFLSLMGLIHKPQVFAIAGA 546
OY 192 PVTLMIFYDGTGYSRRYMGHPDQNEGGYVLGSVAMQAEKFPSEPNRLLILHGFLENVFA 251
DB 547 PVTVMAYDGTGYSRRYMGHPDQNEGGYVLGSVAMQAEKFPSEPNRLLILHGFLENVFA 606
OY 252 HTSILSLFLVAGKRPYDQIY 272
DB 607 HTNFLVSQLIRAGKRYQLQY 627

RESULT 15
O96NT8 PRELIMINARY; PRT; 312 AA.
AC O96NT8;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Hypothetical protein FLJ30094.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuna M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Maeno Y., Nagai K., Isegai T.,
RT "NEO human CDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054656; BAB70784.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR00379; Ser_estr_site.
DR Pfam; PF00326; Peptidase_S9; 1.
KM Hypothetical protein.
SQ SEQUENCE 312 AA; 35518 MW; ABE940AFC5877717 CRC64;

Query Match 55.9%; Score 938.5; DB 4; Length 312;
Best Local Similarity 55.4%; Pred. No. 3.3e-77;
Matches 173; Conservative 38; Mismatches 52; Indels 49; Gaps 1;

OY 38 ISQCHDFISKYNQKNPHCVSLYKLSPPEDPTCKTEKFPATILDSAGPLPDYTPPEIF 97
DB 1 MSQNDMFVSHYSVSTPPCVHVVYLSGPDDDLHKOPRFWASMEASCPDYVPPPIF 60
OY 98 SPESRTGTYGMLYKPHDLOPGKKYPTVFLFYGGP----- 134
DB 61 HFTRSDVRLYGMLYKPHALOPGKHPVTLFLVYGQVQLVNNSEFKIKYLRNLTLASLG 120
OY 135 -----GQIEIDDOVEGLQYLASRYDFIDLRVGIHGSY 168
DB 121 YAVVVIDGRSGCQGLRFEGLXNMQQVEIEDQVEGLQYAEKRYGFDLSRVAIHGWSY 180
OY 169 GGYLSLWLMQSRDIFRYAIGAPVTLMIFYDGTGYSRRYMGHPDQNEGGYVLGSVAMQAE 228
DB 181 GGFLSLMGLIHKPQVFAIACAPVTVMAAYDGTGYSRRYMGHPDQNEGGYVLGSVAMQAE 240
OY 229 KFPSEPNRLLILHGFLENVFAHTSILSLFLVAGKRYDQIYQERHSIRVPSGEHY 288
DB 241 KLPSEPNRLLILHGFLENVFAHTNFLVSQLIRAGKRYQLQIYPNERHSIRCPESGHHY 300
OY 289 ELHLHLHYLOENL 300
DB 301 EVTLHLFLQEYL 312

```

Search completed: October 15, 2003, 17:13:43  
 Job time : 21.8265 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:15:46 (Search time 3384.49 Seconds

(without alignments)  
3747.092 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1680

Sequence: 1 FEGTKDSPLEHLYVSVVN.....HLHLVLENGSRIRALKVI 310

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10070464/runat\_15102003\_113553\_24810/app\_query.fasta\_1.2652  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10070464 @CGN 1.1 13561 @runat 15102003 113553 24810 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG-SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: \*  
1: gb\_ha: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sbs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: gb\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sbs: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_hlg\_hum: \*  
31: em\_hlg\_inv: \*  
32: em\_hlg\_other: \*  
33: em\_hlg\_mus: \*  
34: em\_hlg\_pln: \*  
35: em\_hlg\_rod: \*  
36: em\_hlg\_mam: \*  
37: em\_hlg\_vrt: \*  
38: em\_sy: \*  
39: em\_hlg\_hum: \*  
40: em\_hlg\_mus: \*  
41: em\_hlgc\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1680	100.0	1197	9 AF221635	AF221635 Homo sapi
2	1645.5	97.9	2649	6 AX354795	AX354795 Sequence
3	1645.5	97.9	2656	9 AY172659	AY172659 Homo sapi
4	1645.5	97.9	2671	6 AX608725	AX608725 Sequence
5	1645.5	97.9	3106	6 AX342633	AX342633 Sequence
6	1645.5	97.9	3127	9 AF221634	AF221634 Homo sapi
7	1645.5	97.9	3143	6 AX354793	AX354793 Sequence
8	1645.5	97.9	4535	9 BC040203	BC040203 Homo sapi
9	1644.5	97.9	4676	6 AX608743	AX608743 Sequence
10	1625.5	96.8	4309	6 AX608737	AX608737 Sequence
11	1625.5	96.8	4829	6 AX608735	AX608735 Sequence
12	1605.5	95.6	4799	10 BC043124	BC043124 Mus muscu
13	1407	83.8	3130	9 BC030688	BC030688 Homo sapi
14	1401	83.4	2161	6 BD157001	BD157001 Primer fo
15	1401	83.4	2161	9 AK027826	AK027826 Homo sapi
16	1377.5	82.0	1659	9 AF221636	AF221636 Homo sapi
17	1377.5	82.0	2668	6 AX405771	AX405771 Sequence
18	1377.5	82.0	2778	6 AK000290	AK000290 Homo sapi
19	1377.5	82.0	2842	6 AX405770	AX405770 Sequence
20	1372	81.7	2510	6 AX338497	AX338497 Sequence
21	1350.5	80.9	4685	6 AX608745	AX608745 Sequence
22	1352	80.5	4523	6 AX608731	AX608731 Sequence
23	1093.5	65.1	2546	9 BC000970	BC000970 Homo sapi
24	1093.5	65.1	2602	9 AY172660	AY172660 Homo sapi
25	1093.5	65.1	2617	6 AX608727	AX608727 Sequence
26	1093.5	65.1	2676	6 AX524928	AX524928 Sequence
27	1093.5	65.1	3000	9 AF542510	AF542510 Homo sapi
28	1093.5	65.1	3716	6 AX480934	AX480934 Sequence
29	1093.5	65.1	4219	6 AX608751	AX608751 Sequence
30	1093.5	65.1	4232	9 BC037948	BC037948 Homo sapi
31	1093.5	65.1	4295	9 AF452102	AF452102 Homo sapi
32	1093.5	65.1	4302	6 AX608747	AX608747 Sequence
33	1086.5	64.7	1083	9 AF221637	AF221637 Homo sapi
34	1086.5	64.7	4180	6 AX608759	AX608759 Sequence
35	1086.5	64.7	4263	6 AX608757	AX608757 Sequence
36	1084.5	64.6	2261	6 AX713363	AX713363 Sequence
37	1084.5	64.6	2261	9 AK054656	AK054656 Homo sapi
38	968.5	57.6	3243	9 HSM805448	AK833376 Homo sapi
39	934	55.6	873	6 AX081310	AX081310 Sequence
40	929	55.3	735	6 AX524942	AX524942 Sequence
41	929	55.3	1002	6 AX524939	AX524939 Sequence
42	872.5	51.9	4076	6 AX608755	AX608755 Sequence
43	872.5	51.9	4159	6 AX608753	AX608753 Sequence
44	865.5	51.5	4037	6 AX608763	AX608763 Sequence
45	865.5	51.5	4120	6 AX608761	AX608761 Sequence

RESULT 1

#### ALIGNMENTS

AF221635  
LOCUS AF221635 1197 bp mRNA linear PRI 05-NOV-2000  
DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, partial cds, alternatively spliced.  
ACCESSION AF221635  
VERSION AF221635.1 GI:11095189  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1197)  
AUTHORS Abbott,C.A., Yu,D.M., Woollett,E., Sutherland,G.R., McCaughan,G.W. and Gorrell,M.D.  
TITLE Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homolog, DPP8  
JOURNAL Eur. J. Biochem. 267 (20), 6140-6150 (2000)  
MEDLINE 20467194  
PUBMED 11012666  
REFERENCE 2 (bases 1 to 1197)  
AUTHORS Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUN-2000) A.W. Morrow Gastroenterology and Liver Centre, Centenary Institute of Cell Biology and Cancer Medicine, Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia  
FEATURES  
source location/Qualifiers  
1..1197  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q22"  
/clone="T21"  
/issue\_type="Placenta"  
/clone\_id="lambda triplexXTM: Clontech"  
1..1197  
/gene="DPP8"  
CDS  
1..935  
/gene="DPP8"  
/note="serine protease; member of prolyl oligopeptidase S9b family belonging to enzyme clan SC; additional splice variants of this mRNA are found in GenBank Accession Numbers AF221634, AF221636 and AF221637"  
/codon\_start=3  
/product="dipeptidyl peptidase 8"  
/protein\_id="AAC29767.1"  
/db\_xref="GI:11095190"  
/translation="FEGTKDSPLREHLLYVSVYVNGEVTLLTPDGVSQSCISQCHP FISKYSNQRNHCYSLYKLSPEDDPTCKTEFPAATLLDSAGPLPDTTPPIESFEST TGFLLYGLMYKPHDLPQSKPTVLFYGGQSGIETDDVEGLQYLAIRYDFLDR VGIHGWYGYGLSLMALMQRSDIFRVAIAGAPVLTWTFYDGYTERVWGHDDQEGY YLGSVMAQAEKFPSEPNRLHLHGFLENFAPTSLLSFLVAGKRPYDQLIVPQGR HSIRVPSGEHYELHLHYLOENLGSRAALAKVI"  
BASE COUNT 357 a 258 c 242 g 340 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.47e-156 Length: 1197  
Score: 1680.00 Matches: 310  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-10-070-464-3 (1-310) x AF221635 (1-1197)  
QY 1 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyValValSerTyValAsn 20  
Db 3 TTTAAAGGCACCAAGACTCCCTTTAGAGCATCACTGTAAGTAGTAGTAAAT 62  
QY 21 ProGlyIuValThrArgLeuThrAspArgGlyTySerHisSerCysCysIleSerGln 40  
Db 63 CCTGAGAGGCTGACAAGGCTGACCGTGGCTACTACATTCTTGTCGATCAGTCAG 122

QY 41 HisCysAspPhePheIleSerLysTySerAsnGlnLysAsnProHisCysValSerLeu 60  
Db 123 CACTGTGACTCTTTTAATAGTAAGTAGACAGAAAGATCCAGCTGTGCTCCCTT 182  
QY 61 TyrIleLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 80  
Db 183 TACAACTATCAAGTCCCTGAAGATGACCCAACTTGCAAAAGAAATTTTGGGCCACC 242  
QY 81 IleLeuAspSerLysGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100  
Db 243 ATTTGGATTACAGAGGCTCTCTTCGACTATACCTCCAGAAATTTCTTTGAA 302  
QY 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyLysProHisAspLeuGlnProGly 120  
Db 303 AGTACTACTGATTTTACATTGTATGGATGCTCTCAGACCTCATGATCAGACCTGGA 362  
QY 129 LysLysTyProThrValLeuPheIleTyGlyGlyProGlnGlnIleGluIleAsp 140  
Db 363 AAGAAATATCTACTGCTGCTGCTCATATATGTCGTCCTCAGAGGTCAAAATGAAATTGAC 422  
QY 141 AspGlnValGluGlyLeuGlnTyLysLeuAlaSerArgTyRAspPheIleAspLeuAspArg 160  
Db 423 GATCAGGTGGAAGACCTCCATATCTAGCTTCTGATATATATTCATTGATGATCGT 482  
QY 161 ValGlyIleHisGlyTyrSerTyGlyGlyTyLysSerLeuMetAlaLeuMetGlnArg 180  
Db 483 GTGGGCAATCCACGGCTGTCTCTATGAGGATACCTCTCCCTGATGGCATTAACAGAG 542  
QY 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyRAsp 200  
Db 543 TCAGATATCTTCAGGGTGTGCTATGCTGTCGGGCCCACTCCTGTGATCTTCATGAT 602  
QY 201 ThrGlyTyThrGluArgTyMetGlyHisProAspGlnAsnGlnGlnIleTyTyLysLeu 220  
Db 603 ACAGGATACAGGAACCTTATATGGGTACCTGACAGAAAGAAAGGGCTATTACTTA 662  
QY 221 GlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeu 240  
Db 663 GGATCTGTGGCAATGACAGAGAAAGTCCCTGACCAACCAATCGTTTACTGCTCTTA 722  
QY 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuSerPheLeu 260  
Db 723 CATGGTTTCTCGATGAGTAATGTCATTTGCAATTCAGATATTAATGAGTTTCTTA 782  
QY 261 ValArgAlaGlyLysProTyRAspLeuGlnIleTyRProGlnGlnIleGlnHisSerIleArg 280  
Db 783 GTAGGGCTGGAAGCCATATGATTTACAGTCTATCTCAGAGAGACACAGCATTAAGA 842  
QY 281 ValProGluSerGlyGluHisTyRGluleuHisLeuLeuHisTyRLeuGlnIleAsnLeu 300  
Db 843 GTTCCGTAATCGGAGAACTTATGAACTGATCTTTGGCATCTTCAAGAAACCTT 902  
QY 301 GlySerArgIleAlaIleLeuLysValIle 310  
Db 903 GGATCACTATGTGCTCTTAAAGTGATA 932  
RESULT 2  
AX354795 2649 bp DNA linear PAT 06-FEB-2002  
LOCUS AX354795  
DEFINITION Sequence 3 from Patent WO0179473.  
ACCESSION AX354795  
VERSION AX354795.1 GI:18619528  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Meyers,R.A. and Williamson,M.  
TITLE 21953, a human prolyl oligopeptidase family member and uses thereof  
JOURNAL Patent: WO 0179473-A 3 25-0CF-2001;  
Millennium Pharmaceuticals, Inc. (US)



```

FEATURES
    source
        1..2649
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
BASE COUNT      803 a      514 c      586 g      746 t
ORIGIN
Alignment Scores:
Pred. No.:      9.63e-153      Length:      2649
Score:          1645.50      Matches:      310
Percent Similarity: 86.35%      Conservative: 0
Best Local Similarity: 86.35%      Mismatches: 0
Query Match:    97.95%      Indels:      49
DB:             6      Gaps:      1
US-10-070-464-3 (1-310) x AX354795 (1-2649)

Qy      1  PheGluGlyThrIlyAspSerProLeuGluHisIleuTyValIleSerTyValAsn 20
Db      1570 TTGAAAGGACCAAGACTCCCTTTAGAGCATCAGCTGTAAGTACGTTACGTAAT 1629

Qy      21  ProGlyGluValIThrArgLeuThrAspArgGlyTySerHisSerCysIleSerGln 40
Db      1630 CCGAGAGGAGGACAGGCTGACTGACCGGCTGACTCAGCTTCCTGCTGATCAGTCAG 1689

Qy      41  HisCysAspPhePheIleSerIleTySerAsnGlnIlyAsnProHisCysValSerIleu 60
Db      1690 CACTGTACTCTTTATATAGTAATATATAGTACCAAGAAATCCACTGTGTCTCCCTT 1749

Qy      61  TyrIlySerSerProGluAspAspProThrCysIlyThrIlyGluPheIleAlaThr 80
Db      1750 TACAAAGCATCAAGTCCGAAAGATGACCACTTGCAAAAACAAAGAAATTTGGGCCACCC 1809

Qy      81  IleuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGlu 100
Db      1810 ATTGTGATTCAGAGAGGCTCTTCCTGCACTACTCTCCAGAAATTTCTCTTTTGAA 1869

Qy      101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyIlyPSPProHisAspLeuGlnProGly 120
Db      1870 AGTACTACTGGATTACCTTTATGGATGCTCTTACAAAGCTTCATGATCTACAGCTCGGA 1929

Qy      121 LysIlyTyProThrValIleuPheIleTyGlyIleProGln----- 134
Db      1930 AAGAAATATCTCTACTGTCTGTATATATGTGTCTCTCAGTGACAGTTGGTGAATAT 1989

Qy      134 ----- 134
Db      1990 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTAGCCTCTAGGTTATGTGTT 2049

Qy      134 ----- 134
Db      2050 GTATGATAGACAAACAGGGATCTGTTCACGAGGGCTTAATTTGAAGGGCCTTTAAA 2109

Qy      135 -----GlyGlnIleGluIleAspAspGlnValIleGluIleGluIleAsn 151
Db      2110 TATATAATGGTCAATATAGAAATTCAGCATGAGTGAAGGACTCCCAATATCTAGCTTCT 2169

Qy      152 ArgTyTrpAspPheIleAspLeuAspArgValGlyIleHisGlyTTPSerTyIleGlyIleTy 171
Db      2170 CGATATATATTCATGACTAGATGCTGTGGGCAATCCACGGCTGCTCTTATGAGGATAC 2229

Qy      172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
Db      2230 CTCTCCCTGATGCGATTATATGACAGAGTCAATATCTTCAGGGTGTCTATTTCTGGGACC 2289

Qy      192 ProValThrLeuTrpIlePheTyAspThrGlyIleTyThrGluArgTyIleMetGlyHisPro 211
Db      2290 CCAAGTCACTCTGTGATCTCTATGATATACAGGATACACGGAACGTTATATGGGTCACT 2349

Qy      212 AspGluAsnGluGlnGlyIleTyTrpLeuGlySerValAlaMetGlnAlaGluIlyPhePro 231
Db      2350 GACCAGAAATGAACGGGCTTACTTACTTAGAGATCTGTGGCAATGCAAGAAAGTTTCCCC 2409

```

```

Qy      232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
Db      2410 TCTAACCAATTCGTTTACTGCTCTTACATAGTGTTCCTCGATGAGATGTCATTTTGCA 2469

Qy      252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIlyProTyAspLeuGlnIle 271
Db      2470 CATACAGTATATATTCAGATGTTTATGTAGAGGCTGGAAAGCCATATGATTACAGATC 2529

Qy      272 TyrProGlnIleuArgHisSerIleArgValIProGluSerGlyIleHisGlyIleuHis 291
Db      2530 TATCCTCAGAGAGACACAGCATTAAGTCTCTCAATTCGGAGGAACATTATGAATCAGCAT 2589

Qy      292 LeuLeuHisTyTrpLeuGlnIleuAsnLeuGlySerArgIleAlaIleLeuValIle 310
Db      2590 CTTTGCACTTACTTCAGAAACCTTGATCAGCATATGCTGCTCTAAAGTGATA 2646

RESULT 3
AY172659
LOCUS      2656 bp      mRNA      linear      PRI 08-JAN-2003
DEFINITION Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPP1)
ACCESSION AY172659
VERSION   AY172659.1 GI:27549549
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Bukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
TITLE Novel Serine Protease Genes Related To DPP1V
JOURN. Patent: US (WO 0231134)-A 18-Apr-2002;
REFERENCE
AUTHORS Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2002) Ferring Research Institute, 3550 General
Atomics Ct., San Diego, CA 92121, USA

FEATURES
    source
        1..2656
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /chromosome="15"
        /map="15q22.1-q22.2"
        1..2656
        /gene="DPP1"
        8..2656
        /gene="DPP1"
        /codon_start=1
        /product="dipeptidyl peptidase IV-related protein-1"
        /protein_id="AA017261.1"
        /db_xref="GI:27549550"
        /translation="MAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWGQ
        LKLLADTRKYGHMAKAPHDPMFVGNRPOGDSRIRIYLLAMSGNREUTYLSSEI
        PKTINRAVLMISMKPLLDLFOATLIOGWMSREBELREKRICTGATISVDYQSGG
        TIFPQAGSGIYHKKDGPQFQQPLRPNUVETSCNIRNDPKICPADPMIAFATSHN
        DIVISNIVREERRLTYVHNELANMEDASAGATFVLQEEFDRIYSGYVPMWCPAETT
        PSQGLIKRLIYEENDESEVEI IHTVSPMLTBRADSRFYKPTGRANRKYVFKSEIM1
        DAGRGILRIDVDELIOFELIFEGVEIYIAGWTPBEGVAMSILLDSGRLQIVLIS
        PELFIPVEDDVMRRLIESVPSMPLIIEETTDIWINIHIDIFHVPGSHSEEIF
        IFASECKTGFRHLYKITSLSKSKYRSGLPAPSPKPIKEBIAITGSENVILDR
        HGSNIQVDERRLVYREGTQSLLENHLYVSVNPEVRLTDRGISHCCISQHD
        FLSKYSNQNPCHVSLYKISSPEDPTCTKTFEWMATILDSAGLPDYTPPELIFSES
        TTFPTLYGMLYKPHDLOPKKPYPLVFIYGGPOVOLVNNRFGVYKFRNTLTLISGLY
        VVVIDRGSCHRLKEGAFKRYNGQIEIDQVGLQYLASRYDFIDLDVGHGMSY
        GGYLSLALMORSDIFRVAIAGAPVILMIYDVGTERVYMGHPDQNGGYLSVAMQ
        AEKPEBPNNLLHGFIDENVRHARTSILSLFVAAGKRYDILQIYQENHSIRVPS
        GEHYEHLHLHYLOENSGRIALAKVI"
BASE COUNT      804 a      517 c      588 g      747 t
ORIGIN
Alignment Scores: 0

```

Pred. No.: 9.66e-153 Length: 2656  
 Score: 1645.50 Matches: 310  
 Percent Similarity: 86.35% Conservative: 0  
 Best Local Similarity: 86.35% Mismatches: 0  
 Query Match: 97.95% Indels: 49  
 DB: 9 Gaps: 1

US-10-070-464-3 (1-310) x AY172659 (1-2656)

QY 1 PhegluglyThrIlyAspSerProleugluhshisIeuYrValSerTyValasn 20  
 Db 1577 TTGAAGGACCAAGACTCCCTTTAAGCATCACTGTACGTACGTACCTTAAT 1636  
 QY 21 ProgljgluValThrArgleuThrAspArglyTySerhissSerCyasIlesergln 40  
 Db 1637 CCTGGAGAGTGACAAAGCTGACTGACCGGTGGCTACTCACATTTCTGTGATCATGTCAG 1696  
 QY 41 HisCyasAspPhePheIleSerIlySerAsngInIlysaAnProhIsCyValSerIeu 60  
 Db 1697 CACTGTGACTCTTTATAGTAGTAGTAGTAACCAAGAAATCCACACTGTGTCCCTT 1756  
 QY 61 TyrlIlySerSerProgluAspAspProthrCylsVsthrIlysgluPheTrpAlaThr 80  
 Db 1757 TACAAGCTATCAAGTCTGAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCCACC 1816  
 QY 81 IleuAspSerAlaGlyProleuProAspTyThrProProgluIlePheSerPheglu 100  
 Db 1817 ATTTTGATTCAGACAGGTCTCTCTCTGATCATACTCTCCAGAAATTTCTCTTTTAA 1876  
 QY 101 SerThrThrIlyPheThrIleuTyrgIlyMetIeuTyrlYsPProhIsAspLeugInProgly 120  
 Db 1877 AGTACTACTGATTTACATTTGTATGGAGATGCTTCAACAAAGCTCATGATCTACAGCTTGA 1936  
 QY 121 LysIlyTyPProthrValIeuPheIleTyrgIlyProglIn----- 134  
 Db 1937 AAGAAATATCTCTACTGTCTCTGTTCATATATGTGTCTCTCAGGTGCACTGGTAATAT 1996  
 QY 134 ----- 134  
 Db 1997 CGATTTAAAGAGTCAAGTATTTCCGCTTGATAACCTAGCCCTCTAGTTATGTGTT 2056  
 QY 134 ----- 134  
 Db 2057 GTAGTATAGACAACAGGGGATCCTGTACACGAGGGCTTAATTTGAAGGGCCTTTAA 2116  
 QY 135 -----GlyInIleGluIleAspAspGlnValGluGlyLeuGlnTyrlIleuAlaser 151  
 Db 2117 TATAAAATGGTCAAAATGAATTTGACATCAGGTGGAAAGACTCCAAATATCTAGCTTCT 2176  
 QY 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTPSerTyrgIlyGlyTy 171  
 Db 2177 CGATATGATTTCAATTGACTTGAATCGTGGGCACTCCAGGCTGGTCTATGGAGATAC 2236  
 QY 172 LeuSerIeuMetAlaIeuMetGlnAspSerAspIlePheArgValAlaIleAlaGlyAla 191  
 Db 2237 CTCCTCCGATGGCATTTATATGACAGAGTCAAGATATCTTCAAGGGTGTCTATTTGGGGCC 2296  
 QY 192 ProValThrIeuTrpIlePheTyAspThrIlyTyThrGluArgTyIlyMetGlyVhiAsp 211  
 Db 2297 CCACTCACTCTGTGGATTTCTATGATATCAGAGATACAGAACTTATATGGGTCACTCCT 2356  
 QY 212 AspGlnAsngInGlyTyTyTyIleugIlySerValAlaMetGlnAlaGluIlyAspPhe 231  
 Db 2357 GACAGAGATGAACAGGGCTATTACTTAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCC 2416  
 QY 232 SerGluProAsnAspLeuIleuIlysglyPheIleuAspGluAsnValHisPheAla 251  
 Db 2417 TCTAAACCAATCTTTTACTCTCTTACTAGTGTCTTCTGGATGAGAAATGTCATTTTGCA 2476  
 QY 252 HisThrSerIleIleuIleuSerPheIleuValArgAlaGlyIlyProTyAspLeuGlnIle 271  
 Db 2477 CATACAGATATATATGAGTTTATAGTGGGGCTGGAAGGCAATATGATTTACAGATC 2536

QY 272 TyrProgljgluValThrArgleuThrAspArglyTySerhissSerCyasIlesergln 291  
 Db 2537 TATCTTCAGAGAGACACAGCATTAAGTCTCTGAATCGGAGAACTATTATGAATGATCAT 2596  
 QY 292 LeuIleuIlyTyrlIleugInIlysaAnIlyserArgIleAlaIleuIlyValIle 310  
 Db 2597 CTTTTCAGTACTCTTCAAGAAACCTTGATCAGATATGCTGTCTTAAAGTGATA 2653

RESULT 4  
 AX608725  
 LOCUS AX608725 2671 bp DNA linear PAT 17-FEB-2003  
 DEFINITION Sequence 2 from Patent WO0231134.  
 ACCESSION AX608725  
 VERSION AX608725.1 GI:28404298  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 Q. S., Akinsanya, K. O., Riviere, P. J. and Junien, J. L.  
 Novel serine protease genes related to dpp11  
 Patent: WO 0231134-A 2 18-Apr-2002;  
 Ferring BV (NL)

FEATURES  
 source  
 1..2671  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 805 a 524 c 594 g 748 t  
 ORIGIN

#### Alignment Scores:

Pred. No.: 9.72e-153 Length: 2671  
 Score: 1645.50 Matches: 310  
 Percent Similarity: 86.35% Conservative: 0  
 Best Local Similarity: 86.35% Mismatches: 0  
 Query Match: 97.95% Indels: 49  
 DB: 6 Gaps: 1

US-10-070-464-3 (1-310) x AX608725 (1-2671)

QY 1 PhegluglyThrIlyAspSerProleugluhshisIeuYrValSerTyValasn 20  
 Db 1577 TTGAAGGACCAAGACTCCCTTTAAGCATCACTGTACGTACGTACCTTAAT 1636  
 QY 21 ProgljgluValThrArgleuThrAspArglyTySerhissSerCyasIlesergln 40  
 Db 1637 CCTGGAGAGTGACAAAGCTGACTGACCGGTGGCTACTCACATTTCTGTGATCATGTCAG 1696  
 QY 41 HisCyasAspPhePheIleSerIlySerAsngInIlysaAnProhIsCyValSerIeu 60  
 Db 1697 CACTGTGACTCTTTATAGTAGTAGTAGTAACCAAGAAATCCACACTGTGTCCCTT 1756  
 QY 61 TyrlIlySerSerProgluAspAspProthrCylsVsthrIlysgluPheTrpAlaThr 80  
 Db 1757 TACAAGCTATCAAGTCTGAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCCACC 1816  
 QY 81 IleuAspSerAlaGlyProleuProAspTyThrProProgluIlePheSerPheglu 100  
 Db 1817 ATTTTGATTCAGACAGGTCTCTCTCTGATCATACTCTCCAGAAATTTCTCTTTTGA 1876  
 QY 101 SerThrThrIlyPheThrIleuTyrgIlyMetIeuTyrlYsPProhIsAspLeugInProgly 120  
 Db 1877 AGTACTACTGATTTTATATGATGTATGGATGCTCTACAAAGCTCATGATCTACAGCTTGA 1936  
 QY 121 LysIlyTyPProthrValIeuPheIleTyrgIlyProglIn----- 134  
 Db 1937 AAGAAATATCTCTACTGTGTCTCATATATGTGTGTCTCAGGTGCACTGGTAATAT 1996  
 QY 134 ----- 134  
 Db 1997 CGATTTAAAGAGTCAAGTATTTCCGCTTGATAACCTAGCCCTCTAGGTTATGTGTT 2056

```

QY      134 ----- 134
DB      2057 GTAGTGATAGACAAACAGGGGATCTGTACCGAGGCTTAATTTGAAAGCGCTTTAA 2116
QY      135 -----GlyGlnIleGluIleAspArgIleValGluIleGlnIleuValAsn 151
DB      2117 TATAAATATGGTCAAAATAGAAATTTGACGATCAGGTGGAAGGATCCCAATATCTAGCTTCT 2176
QY      152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 171
DB      2177 CGATATGATTTTCAATGACTTAGATCGGTGGCATCCACGGCTGGTCTTATGAGAGATAC 2236
QY      172 LeuSerIleuMetAlaIleuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB      2237 CTCTCCCTGATGGCATTTAATGACAGAGTCAGATATCTTCAGGGTTGCTATGCTGGGGCC 2296
QY      192 ProValThrIleuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
DB      2297 CCAGTCACTCTGTGATCTTCTTATGATACAGATACCGGAACTTATATGGGTCAACCT 2356
QY      212 AspGlnAspGlnGluIleGlyTyrTyrIleuGlySerValAlaMetGlnAlaGlyHisPhePro 231
DB      2357 GACCAAGATGAAACAGGGCTATTACTTAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCC 2416
QY      232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
DB      2417 TCTGAACCAATTCGTTTACTGCTCTTACATGCTTCTCTGATGAGAAATGTCATTTTGA 2476
QY      252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
DB      2477 CATACAGATATATTAATGATGATTTTATGAGGGCTGGAAGCCATATGATTTACAGATC 2536
QY      272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291
DB      2537 TATCTCAGAGAGACACAGCATTAAGAGTTCCTGAATCGGGAACCATTAATGACGCAAT 2596
QY      292 LeuLeuHisTyrIleuGlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 310
DB      2597 CTTTTCACACTACCTTCAAGAAACCTTGGATCAAGATTCCTCTCTTAATAAGTATTA 2653

RESULT 5
AX342633
LOCUS      AX342633          3106 bp      DNA      linear      PAT 12-JAN-2002
DEFINITION Sequence 30 from Patent WO0198468.
ACCESSION  AX342633
VERSION     AX342633.1  GI:18152030
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Yue,H., Elliott,V.S., Gandhi,A.R., Lal,P., Au-Young,J.,
            Tribouley,C.M., Deleage,A.M., Baughn,M.R., Nguyen,D.B., Lee,E.A.,
            Hatfield,A., Khan,F.A., Wallis,N.K., Yao,M.G., Lu,D.A., Patterson,C.,
            Tang,Y.T., Walsh,R.T., Azimzai,Y., Rankumar,J., Xu,Y. and Reddy,R.
            Patent: WO 0198468-A 30 27-DEC-2001;
            Incyte Genomics, Inc. (US)
FEATURES
            Location/Qualifiers
                source          1..3106
                                /organism="Homo sapiens"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:9606"
                                /note="Incyte ID No: 7160544CB1"
BASE COUNT  928 a      633 c      704 g      841 t
ORIGIN

```

## Alignment Scores:

```

Pred. No.:      1,16e-152      Length:      3106
Score:          1645.50        Matches:      310
Percent Similarity: 86.358      Conservative: 0
Best Local Similarity: 86.358      Mismatches: 0

```

```

Query Match:      97.95%      Indels:      49
DB:               6          Gaps:      1
US-10-070-464-3 (1-310) x AX342633 (1-3106)
QY      1 PheGluGlyThrLysAspSerProLeuGluHisIleuTyrValIleSerTyrValAsn 20
DB      1772 TTTGAAAGGACCAAGAGCTCCCTTTAGAGCATCACCTGATGACATGACATGAT 1831
QY      21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysValIleSerGln 40
DB      1832 CCTGGAGAGGTGCAAGAGCTGACTGACCGGTGCTACATCTTGGCTCATCAGTCAAG 1891
QY      41 HisCysAspPhePheIleSerLysTyrSerAspGlnLysAsnProHisCysValIleSerLeu 60
DB      1892 CACTGGAGCTCTTTATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 1951
QY      61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 80
DB      1952 TACAGCTATCAAGATCCTGAGATGACCCCACTTGCAAAACAAAGGATTTTGGCCACC 2011
QY      81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100
DB      2012 ATTTTGATTCACAGAGCTCTCTTCTGACTATACCTCCAGAAATTTTCTCTTTTGA 2071
QY      101 SerThrThrGlyPheThrIleuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
DB      2072 AGTACTACTGAGATTTCATGTATGAGATGCTTACAGGCTCATGATGATACAGCTTGA 2131
QY      121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
DB      2132 AAGAAATATCCTACTGCTGCTTGCATATATGATGGTCTCTCAGGTGACGTTGGAATAT 2191
QY      134 ----- 134
DB      2192 CGGTTTAAAGAGTCAAGTATTTCCGCTTAATACCTAGCCTCTAGGTATATGCTT 2251
QY      134 ----- 134
DB      2252 GTAGTGATAGACAAACAGGGGATCTGTACCGAGGCTTAATTTGAAAGCGCTTTAA 2311
QY      135 -----GlyGlnIleGluIleAspArgIleValGluIleGlnIleuValAsn 151
DB      2312 TATAAATATGGTCAAAATAGAAATTTGACGATCAGGTGGAAGGATCCCAATATCTAGCTTCT 2371
QY      152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 171
DB      2372 CGATATGATTTTCAATGACTTAGATCGGTGGCATCCACGGCTGGTCTTATGAGAGATAC 2431
QY      172 LeuSerIleuMetAlaIleuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB      2432 CTCTCCCTGATGGCATTTAATGACAGAGTCAGATATCTTCAGGGTTGCTATGCTGGGGCC 2491
QY      192 ProValThrIleuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
DB      2492 CCAGTCACTCTGTGATCTTCTATGATACAGATACACGAACTTATATGGGTCAACCT 2551
QY      212 AspGlnAspGlnGluIleGlyTyrTyrIleuGlySerValAlaMetGlnAlaGlyHisPhePro 231
DB      2552 GACCAAGATGAAACAGGGCTATTACTTAGATCTGTGGCCATCAGCAAGAAAGTTCCCC 2611
QY      292 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
DB      2612 TCTGAACCAATTCGTTTACTGCTTACATGGTTCCTGAGATGAGAAATTCATTTTGA 2671
QY      252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
DB      2672 CATACAGATATATTAATGATGATTTTATGAGAGGCTGGAAGCCATATATTTACAGATC 2731
QY      272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291
DB      2732 TATCTCAGAGAGACACAGCATTAAGAGTTCCTGATCGGAGAACATTAATGAACTGCAAT 2791

```

QY 292 LeuLeuHisTyrLeuGlnGluLeuGlySerArgIleAlaLeuValIle 310  
 DB 2792 CTTTGACACTACCTTCAAGAAACCTTGATCAGTATGCTGCTTAAAGTGATA 2848

RESULT 6  
 AF221634  
 LOCUS AF221634 3127 bp mRNA linear PRI 05-NOV-2000  
 DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.  
 ACCESSION AF221634  
 VERSION AF221634.1 GI:11095187  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 3127)  
 AUTHORS Abbott,C.A., Yu,D.M., Woolf, E., Sutherland,G.R., McCaughan,G.W.  
 and Gorell,M.D.  
 TITLE Cloning, expression and chromosomal localization of a novel human  
 dipeptidyl peptidase (DPP) IV homolog, DPP8  
 JOURNAL Eur. J. Biochem. 267 (20), 6140-6150 (2000)  
 MEDLINE 20467194  
 PUBMED 11012666  
 REFERENCE 2 (bases 1 to 3127)  
 AUTHORS Abbott,C.A., Yu,D.M., McCaughan,G.W. and Gorell,M.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUN-2000) A.W. Morrow Gastroenterology and Liver  
 Centre, Centenary Institute of Cell Biology and Cancer Medicine,  
 Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia  
 FEATURES  
 Source  
 1. .3127  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="15"  
 /map="15q22"  
 1. .3127  
 /gene="DPP8"  
 214. .2862  
 /gene="DPP8"  
 /note="serine protease; member of prolyl oligopeptidase 89b  
 family belonging to enzyme clan SC; alternatively spliced  
 forms of this mRNA are found in AF221635, AF221636 and  
 AF221637"  
 /product="dipeptidyl peptidase 8"  
 /codon\_start=1  
 /protein\_id="AAC29766.1"  
 /db\_xref="GI:11095188"  
 /translation="MAAAMEEQLGVEIFETADCEENIESQDRKLEPFYVERYSMSQ  
 LKULADTRKHGIMAKAPHDFFVRNDPQPHSDRIITYLMSGNRENTLFSSEI  
 PKTINRAVALMLSWKPLDLFQATLDYGMYSREBELRERKRIGTVAGIADYHQSG  
 TPLFOAGSGIYHVGDGPGFTQOPLRPMLNVESTCPRIKMDPKICPAIDPMWIAFIHSN  
 DIMINSIVTREERELTYVHNELAMMEDASGATVLEDEEPRFSGYGMCPKRAET  
 PSGGKIRLILEENDESEVEIHYTSPMLSTRADSRFRYPKGTANPKTPKSEIMI  
 DAGSRITDVIDKEILOPELIFEGVEYIAAGRTBPKYKMSILDRSQRLOVLIS  
 PELPIVEDVMEKRLIESVPSVTPPLIIEETDWINIHDFHVPQSHHEEELF  
 IFASECKTGRRLHYKITSILKESKYSKSSGSLPAPSDPKPIKEZLAITGEBEVLGR  
 HGSNIQVDEVRRLVYFEGTDSPLRHLNLYVSVYVPEVRLIDRGYSHSCCSOHD  
 FFSKYSNOKNPKCVSLYKLSPEDEPTCKEPMATILDSAGLPYTPPEIFSFES  
 TTGFTLYGMLYKPHDLOGKKYPTVLFITYGSPQVLYNNRFGKYVRLNTLSLIGV  
 VVVINDNGSCHRGKFGKAPKVKGOIEIDDOVGLQYLAASRDFIDLDYGHGWSY  
 GYLISLALMQRSDIFRVALAGAPVTIMTYIDGTGYTRVNGHPDNEOGYVDSVAHQ  
 AEKPSPEPNRLLLHGFLDENVHFPAHTSILLSPVRAKPYDLOIYFQERHRSRVES  
 GEHYELHLHYLDENLSRIALVKVI"

BASE COUNT 943 a 637 c 706 g 841 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,17e-152 Length: 3127  
 Score: 1645.50 Matches: 310  
 Percent Similarity: 86.35% Conservative: 0  
 Best Local Similarity: 86.35% Mismatches: 0  
 Query Match: 97.95% Indels: 49

DB: 0 9 Gaps: 1  
 US-10-070-464-3 (1-310) x AF221634 (1-3127)

QY 1 PheGluGlyThrTyrAspSerProLeuGluHisIleuTyrValIleSerTyrValAsn 20  
 DB 1783 TTTGAAGGACACCAAGCTCTCCCTTAGACATACACTGATGACGTATGCTAAAT 1842

QY 21 ProGlyValuValThrArgLeuThrAspArgGlyTyrSerHisSerCysSileSerGln 40  
 DB 1843 CTGGAGAGGTGACCAAGGCTGACTGACCGGCTACTCATCTTCTGCTCATCAGCAG 1902

QY 41 HisCysAspPhePheIleSerIleTyrSerAsnGlnLysAsnProHisCysValSerLeu 60  
 DB 1903 CACTGTGACTCTTTATTAAGTATAGTATAGTACAGAAATCCACACTGTGTCTCTT 1962

QY 61 TyrIleuSerSerProGluAspAspProThrCysIleThrIleGluPheTyrPalaThr 80  
 DB 1963 TACAACTATCAAGTCTGTGAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCTACC 2022

QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
 DB 2023 ATTTGGATTGACGAGGTCTCTTCTGACTATATCTCTCAGAAATTTCTCTTTGAA 2082

QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeuGlnProGly 120  
 DB 2083 AGTACTACTGATTTACATTGATGATGAGATGCTCTACAAAGCTCATGATCTACAGCTGGA 2142

QY 121 LysIleTyrProThrValLeuPheIleTyrGlyGlyProGln 134  
 DB 2143 AAGAAATATCTCTACTGCTGCTCATATATATGATGATGATGATGATGATGATGAT 2202

QY 134 134  
 DB 2203 CGGTTTAAAGAGTCAAGTATTTCCGTTGAATACCTACCTCTAGGTATGTGTT 2262

QY 134 134  
 DB 2263 GTAGTGTAGACAAACAGGAGCTCTGTACCGAGGCTTAATTTGAAGGCGCTTTAAA 2322

QY 135 135  
 DB 2323 TATTAATGAGGCTCAATATGAAATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2382

QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 171  
 DB 2383 CAGTATGATTTCACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2442

QY 172 LeuSerIleuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
 DB 2443 CTCTCCCTGATGGCATTAAATGACAGAGGTGATATATCTTCAGGGTTGCTATGCTGGGGCC 2502

QY 192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrCysIleGlyTyrMetGlyHisPro 211  
 DB 2503 CCACTCACTCTGTGATCTTCTATGATACAGATACAGATACAGATACAGATACAGATACAGAT 2562

QY 212 AspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231  
 DB 2563 GACCAAGATGAACAGGCTATTAATCTTAGATCTGTGGCCATGCAAGCAGAAAGTTCTCCC 2622

QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluValIleHisPheAla 251  
 DB 2623 TCTGAACCAAAATGTTTACTGCTCTTACATGATGATGATGATGATGATGATGATGATGAT 2682

QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
 DB 2683 CATACCAATATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2742

QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291  
 DB 2743 TATCCCTCAGAGAGACACAGCATTAAGATGCTCGAATCGGAGAAACATTATGAATGATCAT 2802

QY 292 LeuLeuHisTyrLeuGlnGluLeuGlySerArgIleAlaLeuValIle 310

Db 2803 CTTTGACACTACCTTCAAGAAAACCTTGATCAGGATTCGCTCTTAAGTGATA 2859  
|||||  
RESULT 7  
AX354793 3143 bp DNA linear PAT 06-FEB-2002  
LOCUS Sequence 1 from Patent WO0119473.  
DEFINITION AX354793  
ACCESSION AX354793  
VERSION AX354793.1 GI:18619526  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 Meyers, R.A. and Williamson, M.  
21993, a human prollyl oligopeptidase family member and uses thereof  
Patent: WO 01/9473-A 1 25-OCT-2001;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source  
1. 3143  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
229. 2877  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD22990.1"  
/db\_xref="GI:18619527"  
CDS  
translation="MAAMETSQLGVEIFETADCEINIESQDRPLEPFYVERYSQ  
LKLADLRKHYHMAAKAPHDPMFVKRNDPGPSDRIYVLAAMGKRENTLFSYSEI  
PTIRAAVILMLSMKPLDLFOATLDYGMYSREELLEKRIIGVGLASVYHGGSG  
FTLFOAGSIVYHKGQPGQFOOPLRPLVETSCPNIRMDKLPADDMWAFHSN  
DIVISNITRBERRLTYVNEILANMEBARSAGVATPULOEEFDPSCYCMKPKAT  
PSGKITILVYENDESEVEITIHVTSPLMETRADSPRYPKTKTANPKYTRMSGIMI  
DAEGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGRKAMSLILDRQTRQIVLIS  
PELFIPEDDVWERQLIESVDSVTPILYEETDIDMINHIDIFHVPSHSEIER  
IFASCKTGFRHLKYITSLIKESKYRSQGLPAPSDKPCPKIEIAITSGHEVLEIF  
HGSNIQVDEVRLLVYEGTKDPSLEHILVYVAVGEVTRLTDRGSHSCCISQHC  
FPIKSYNOKNPKCVSLYKLSPEDDPKCTKEPMATLIDAGPLPDYTPREPSFES  
TTGFTLYGMLKYPHLOPKKPYPLFIYGGQVQVLYNNRFGVKYRPLANTLASGIV  
VVVINDRSCSHKGLFEGAFKYMQIEIDDOVEGLQYLASRYDIDIDRVGIIHMSY  
GGYLSLMLMORSDIFRVAIAGAPVLMIFDTGYTERYMGHDPDNEQGYLGSVAMQ  
AAKPSSENNRLILHGLFDENYFAHTILLSFLVRAKPKYDLQIYPOERSHVRPES  
GEHYELHLHYLENNLNGSRIALKVI"  
BASE COUNT 943 a 644 c 712 g 844 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.18e-152 Length: 3143  
Score: 1645.50 Matches: 310  
Percent Similarity: 86.35% Conservative: 0  
Best Local Similarity: 86.35% Mismatches: 0  
Query Match: 97.95% Indels: 49  
Gaps: 1  
US-10-070-464-3 (1-310) x AX354793 (1-3143)  
QY 1 PheGluGlyThrIlyAspSerProLeuGluHsiHleuTyValValSerTyValAsn 20  
Db 1798 TTGGAAGCACCACCAAGACTCCCTTTAGGATCAGCTAGTGAAGTAACTAAT 1857  
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTySerHisSerCysCysIleSerGln 40  
Db 1858 CCGTAGAGGTGACAAAGCTGACTGACCGTGCTACTCATTCTTGCTGATCAGTGACG 1917  
QY 41 HisCysAspPhePheIleSerIlyTySerAsnGlnIlyAsnProHisCysValSerLeu 60  
Db 1918 CACTGTGACTCTTATATAAGTAGATAGTAAACGAGAATCCACACTGTGTGCTCTT 1977  
QY 61 TyrlsIleuSerSerProGluAspAspProThrCysIlyThrIlyGluIlePheTrpAlaThr 80  
Db 1978 TACAAGTATCAAGCTCTGAAGATGACCACTTGCAAAAACAAAGAAATTTTGGCCACC 2037

QY 81 IleLeuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGlu 100  
Db 2038 ATTTGGATTACAGACAGCTCTTCTTCTACTACTCTCTCCAGAAATTTCTCTTTGAA 2097  
QY 101 SerThrThrGlyPheThrIleuTyGlyMetLeuTyIlyPheProHisAspLeuGlnProGly 120  
Db 2098 AGTACTACTGATGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2157  
QY 121 TyrlsIlyTyProThrValLeuPheIleTyGlyGlyProGln----- 134  
Db 2158 AAGAAATATCCACTGTGCTGTTCATATATGATGATGATGATGATGATGATGATGATGAT 2217  
QY 134 ----- 134  
Db 2218 CGGTTAAAGAGTCAAGATTTCCGCTTGAATACCCTAGCTCTAGCTATGCTT 2277  
QY 134 ----- 134  
Db 2278 GTAGTATAGACACAGGGGATCCTGACCGAGGGCTTAATTGAAAGCGCCTTAA 2337  
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyIleuAsp 151  
Db 2338 TATAAATGGGTCAAAATGAAATTCAGATCAGTGGGAAAGGACTCCAATATCTAGCTTCT 2397  
QY 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyGlyGlyTy 171  
Db 2398 CGATATGATTCATGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2457  
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
Db 2458 CTCTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2517  
QY 192 ProValThrLeuTrpIlePheTyAspThrGlyTyThrGluArgTyMetGlyHisPro 211  
Db 2518 CCAGTCACTGCTGATCTTCTATATATACAGATACAGGAACTGATATATGATGATGATGAT 2577  
QY 212 AspGlnAsnGluGlnGlyTyTyIleuGlySerValAlaMetGlnAlaGluIlyAspPhePro 231  
Db 2578 GACCAATATGACAGGCTATTTACTTAGATGATGATGATGATGATGATGATGATGATGATGAT 2637  
QY 232 SerGluProAsnArgLeuLeuLeuLeuLeuIleGlyPheLeuAspGluAsnValHisPheAla 251  
Db 2638 TCTGAACCAATCGTTTACTGCTCTTACTAGGTTTCTCGATGAGAATGATTCATTTTGA 2697  
QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIlyTyTyAspLeuGlnIle 271  
Db 2698 CATACCACTATATTACTAGTTTCTTACTGAGGCTGGAAGCCATATGATTTACAGATC 2757  
QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnIlyTyGlyLeuHis 291  
Db 2758 TATCTCAGAGAGACACAGATTAAGATTCCTGTAATGGGAGAACATTTATGAACCTGAT 2817  
QY 292 LeuLeuHisTyIleuGlnGluAsnLeuGlySerArgIleAlaAlaLeuIlyValIle 310  
Db 2818 CTTTGACACTACCTTCAAGAAAACCTTGATCAGGATTCGCTCTTAAGTGATA 2874  
RESULT 8  
BC040203 4535 bp mRNA linear PRI 03-DEC-2002  
LOCUS Homo sapiens, similar to dipeptidyl peptidase 8, clone MGC:42570  
DEFINITION IMAGE:4824813, mRNA, complete cds.  
ACCESSION BC040203  
VERSION BC040203.1 GI:26007916  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (Bases 1 to 4535)  
Strausberg, R.  
Direct Submission  
Submitted (27-NOV-2002) National Institutes of Health, Mammalian

REMARK  
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Miklos Palakovic, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnarsson, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louie, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

## FEATURES

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IPAK Plate: 71 Row: K Column: 23.

## SOURCE

1. 4535  
Location/Qualifiers

## CDS

```
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="MGC:42570 IMAGE:482413"  
/tissue_type="Testis"  
/clone_id="NH MGC_97"  
/lab_host="DH10B"  
/note="Vector: pBluescript"  
/codon_start=1  
/product="similar to dipeptidylpeptidase 8"  
/protein_id="AAH40203.1"  
/db_xref="GI:2607917"  
/db_xref="LOCUSID:54878"  
/translation="MAAAMETEQLVEIFETADCEINIESODRPLEPFYVERYSWQ  
LKLADTRKYHYMMAKAPHDPMFVKENDPDGPHSDRIYLAAMSGENRENTLPSYSEI  
PKTIRAAVLMISMKPLDLFOATLDYVREBELLEERKIGVGIASDYQVSG  
TFLPQAGSIVHVKVQGGFOQPLRNVLVETSCPNRMDPKCPADPMIAFVHGN  
DIWISNITRREBRLLTYVHNLAMNEEDASAGVATFLQSEPRYSYMMCPAETT  
PSGKITLIILENDESEVEIHTVSPMLERRADSPYRTGTANPVTYRKSEINI  
DAEGRIDVIDRELQPELIPBGEVITARGMTPBEGTAMSLIDRSQTLQIVLIS  
PELIPVEDDVMERQLIESVDSVPLIYEETDIAINIHDIHFVPSQHEEIER  
IFASCKTGFPHLYKITSILKESKYKSSGGLPAPSDKCPIKEBIAITSGEWEILGR  
HGSNIQVDEVRLLVYFEGTKDSPLEHLYVVSYNVGEVTLTDGYSHSCISQCHD  
FEISKYSNOKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDVTPPEISFES  
TTGRTLYGMLYKPHDLQPGKPYTLPIYGGPOVQVUNNRKQKYPRLNTLASIGYV  
VVVIDNRGSCRGKLFEGAFYKMGQIIDDVBEGLQYLAIRYDTIDDRGIGHMSY  
GGYLSLMLMORSDIFRVALGAPVTLMI FYDTGYTERYMGHPDQNEGYLVGVAQO  
AEKPESEPNRLILHGFLENVHFAHTSILSLFLVRACKPYDLQIYPOERSHIVPES  
GEHYEHLHLHLEOENSGRIAAAKVI"
```

BASE COUNT 1313 a 918 c 946 g 1358 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.81e-152 Length: 4535  
Score: 1645.50 Matches: 310  
Percent Similarity: 86.35% Conservative: 0  
Best Local Similarity: 96.35% Mismatches: 49  
Query Match: 97.95% Indels: 1  
DB: 9 Gaps: 1

US-10-070-464-3 (1-310) x BC040203 (1-4535)

Qy 1 PheGluGluThrIleAspSerProLeuGluHisIleLeuTyValValIleSerTyValAsn 20  
Db 3198 TTGAGGACCAAGACTCCCTTTAGGATCCCTGTACGTAGTACGTTACCTAAT 3257  
Qy 21 ProGluValIleThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40  
|||||

Db 3258 CTTGAGAGGTGACAGAGGCTGATGACCGTGGCTTACATCTTCTGCTGATCAGTCA 3317  
Qy 41 HisCysAspPhePheIleSerLeuTySerAsnGlnIleAspProHisCysValSerLeu 60  
Db 3318 CACTGTGACTTCTTTATAGATAGTATAGTACAGAGAAATCCACACTGTGTGCTT 3377  
Qy 61 TyrIleAspSerSerProGluAspAspProThrCysIleThrIleGluPheTrpAlaThr 80  
Db 3378 TACAAGCATCATCAATCTCGAAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACC 3437  
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGlu 100  
Db 3438 ATTTGATTCAGAGGCTCTTCTGCTGATATCTCTCCAGAAATTTCTCTTTTAA 3497  
Qy 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyIleProHisAspLeuGlnProGly 120  
Db 3498 AGTACTACTGATTTACTATGATGTATAGGATGCTTCAAGAGCTCATGATCTACAGCTTGA 3557  
Qy 121 LysIleTyProThrValLeuPheIleTyGlyIleProGln----- 134  
Db 3558 AAGAAATATCTACTGTGCTGCTCATATATGTGTGCTCTCAGTGCACTGTGTAATAT 3617  
Qy 134 ----- 134  
Db 3618 CGTTTAAAGATCAAGTATTTCCCTGTAATCCCTAGCTCTATGTTATGTGTT 3677  
Qy 134 ----- 134  
Db 3678 GTAGTATAGACAAAGGCGATCTGTGACCGAGGCGCTTAATTTGAAGGCGCTTTAA 3737  
Qy 135 -----GlyGlnIleGluIleAspAspGlnValGluIleLeuGlnTyrLeuAlaSer 151  
Db 3738 TATAAATGGGTCAAAATGAAATGACATCAGTGTGAGGACATCCATATCTGACTTCT 3797  
Qy 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyGlyGlyTyr 171  
Db 3798 CGATATGATTTCTATGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 3857  
Qy 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
Db 3858 CTCCTCCGATGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 3917  
Qy 192 ProValThrLeuThrPhePheTyAspThrGlyTyrThrGluAlaGlyTyrMetGlyHisPro 211  
Db 3918 CCAGTCACTCTGTGATCTTCTATGATGATGATGATGATGATGATGATGATGATGATG 3977  
Qy 212 AspGlnAsnGlnGlnGlyTyrTyLeuGlySerValAlaMetGlnAlaGluLeuPhePro 231  
Db 3978 GACCAAGATGAACAGGCTATTTACTTATGATCTGTGCGCATGCAAGCAAGAAAGTTCC 4037  
Qy 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
Db 4038 TCTGAACCAATCGTTTACTCTCTTACATGATGTTCTCGATGAGAAATGCTCATTTTGA 4097  
Qy 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIleGlyIleProTyAspLeuGlnIle 271  
Db 4098 CATACCAATATATTAATGATGATTTTATGATGAGGCTGAGAAACCATATGATTTACAGATC 4157  
Qy 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisIleSerGlyLeuHis 291  
Db 4158 TATCTCAGAGAGACACAGATTAAGACTTCTGTAATCCGAGAGAACTATTAAGACTGAT 4217  
Qy 292 LeuLeuHisTyLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuValIle 310  
Db 4218 CTTTTCACATCACTTCAAGAAACCTTGATGATGATGATGATGATGATGATGATGATGATG 4274  
RESULT 9  
LOCUS AX608743 4676 bp DNA linear PAT 17-FEB-2003  
DEFINITION Sequence 20 from Patent WO0231134.  
ACCESSION AX608743  
VERSION AX608743.1 GI:28404307  
KEYWORDS

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L. Novel serine protease genes related to dprlv Patent: WO 02/31134-A 20 18-Apr-2002;
AUTHORS	Ferrling BV (NL)
TITLE	Location/Qualifiers
JOURNAL	1..4676 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
FEATURES	source
BASE COUNT	1424 a 859 c 979 g 1414 t
ORIGIN	
Alignment Scores:	
Pred. No.:	2,35e-152 Length: 4676
Score:	1644.50 Matches: 308
Percent Similarity:	99.35% Conservative: 0
Best Local Similarity:	99.35% Mismatches: 1
Query Match:	97.89% Indels: 2
DB:	Gaps: 1
US-10-070-464-3 (1-310) x AX608743 (1-4676)	
QY	1 PheGluGIyThrLysAspSerProLeuGIuhShiSLeuTYrValIalSerTYrValAan 20
Dd	1783 TTtGAAGCACCMAAGACTCCCTTTAAGCATCACTGTACGTAGTACGTAAAT 1842
QY	21 ProGIyGIuValThrArgLeuThrAspArgLyTySerHisSerCySeilSerGIin 40
Dd	1843 CCTGAGAGGGAGACAAGCGTCAGTACCGGTGGCTACTCACATTCTTGCGCATCAGACAG 1902
QY	41 HisCyaspPhePheIlleSerLysTrpSerAnGIuLysAspProHisCyValSerLeu 60
Dd	1903 CACTGTGACTCTTTATAGTAGTAGTAAGTAAGTAACAGAAGAACCACTGTGTGCCCTT 1962
QY	61 TyrILysLeuSerSerProGIuAspAspProThrCySLySThrILysGIuPheTrPAIathr 80
Dd	1963 TACAAGCTATCAAGCCCTGAAGATGATCAACCACTGCACAAAACAAAGAAATTTGGGCCACC 2022
QY	81 IleLeuAspSerAlaGIyProLeuProAspTYrThrProProGIuIIlPheSerPheGIu 100
Dd	2023 ATTTTGATTCAGCAGGTGCTCTTCTGACTATACTCCTCCAGAAATTTCTCTTTGAA 2082
QY	101 SerThrThrGIyPheThrLeuTYrGIymetLeuTYrLysProHisAspLeuGIinProGIy 120
Dd	2083 AGTACTACTGCAATTTACATTGTATGGAGTGTCTACAAAGCCTCATGATCTACAGCCTAGGA 2142
QY	121 LysLYSTyrProThrValLeuPheIlleTYrGIyGIyProGIingIyGINileGIuIIleAsp 140
Dd	2143 AAGAAATATCTTACTCTGCTGCTTCTATATATGTGTGC---GGTCATAATGAATAATTTGAC 2198
QY	141 AspGIinValGIyGIyLeuGIinTYrLeuAlaSerArgTYrAspPheIIleAspLeuAspArg 160
Dd	2199 GATCAGGAGGAAGGACTCCAATATCTAGCTTCTCGATATGATTTCAATTGACTTATGATCT 2258
QY	161 ValGIyIIleHisGIyTPRSetTYrGIyGIyTYrLeuSerLeuMetCALaleuMerGIinArg 180
Dd	2259 GTGGGCAATCCACGGCTGTGCTCTATGAGAGAACTCCTCCCGATGGGCAATTATACAGAG 2318
QY	181 SerAspIIlePheArgValAlaIIleAlaGIyAlaProValThrLeuTPRIlePheTYrAsp 200
Dd	2319 TCAGATATCTTCAAGGTGTCATATTCCTGAGGCCCAAGTCACTCTGTGATTTTATAGAT 2378
QY	201 ThrGIyTYrThrGIuArgTYrMetGIyHisProAspGIinAsngINGIyTYrTYrLeu 220
Dd	2379 ACAGAGATACAGGAACGTTATATAGGTGATCCCTGACCAAGATGAACAGGGCTATTACTTA 2438
QY	221 GlySerValAlaMetGIinAlaGIuLyllyPheProSerGIuProAsnArgLeuLeuLeu 240

Dp	2439	GGATCTGTGGCGCATGCAAGCAAGAAAGTTCCTCCCTGACCAAAATCGTTTACTGCTCTTA	2498
Qy	2481	HisGlyPheLeuAapGluAenValHisPheAlaHisThrSerIleLeuLeuSerPheLeu	260
Dp	2499	CATGGTTTCCGTGATGAGAAATGTCATTGTCATGACATACCGATATATTACGAGCTTTT	2558
Qy	261	ValArgAlaGlyIleProTyrAspLeuGlnIleTyrProGlnGluAaGHisSerIleArg	280
Dp	2559	GTCAGGGCTGGAAAGCCATATGATGATTTACAGATCTATCTCCAGAGACACAGCATAGA	2618
Qy	281	ValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAaLeu	300
Dp	2619	GTTCCCTGATCGGAGAACATTATGAACCTGCATTTTGGCACTACCTTCACAGAAACCTT	2678
Qy	301	GlySerArgIleAlaIleAlaLeuLysValIle	310
Dp	2679	GGATCAGGTATGCTGCTCTTAAAGTGATA	2708
RESULT 10			
AA608737			
LOCUS	AX608737	4309 bp	DNA
DEFINITION	Sequence 14 from Patent WO0211134.		linear
ACCESSION	AX608737		
VERSION	AX608737.1	GI:28404304	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	Oi,S., Aktsuanyu,K.O., Riviere,P.J. and Junin,J.L.		
	Novel serine protease genes related to dppiv		
	Patent: WO 0231134-A 14 18-APR-2002;		
	Ferring BV (NL)		
FEATURES			
source	location/Qualifiers		
	1..4309		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	1304 a 790 c 907 g 1308 t		
ORIGIN	0		
	1		
Alignment Scores:			
Pred. No.:	1.62e-150	Length:	4309
Score:	1625.50	Matches:	310
Percent Similarity:	85.87%	Conservative:	0
Best Local Similarity:	85.87%	Mismatches:	0
Query Match:	96.76%	Indels:	51
DB:	6	Gaps:	1
US-10-070-464-3 (1-310) x AX608737 (1-4309)			
Qy	1	PheGluGlyThrIleAspSerProLeuGluHisLeuTyrValValSerTyrValAsn	20
Dp	1263	TTTGAAAGGCACCAAGACTCCCTTTAGACATCACCTGTAGCACTGACGTACGTACGTAAT	1322
Qy	21	ProGlyGluValAlaThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln	40
Dp	1323	CTGTGAGAGGTCAGAACGCTGACTGACCGGTGGCTACTCACATTCCTGGCTGCATCAGTCAG	1382
Qy	41	HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAspProHisCysValSerLeu	60
Dp	1383	CACGTGACTCTTTAATAAGTATAGTATACCAAGAAATCCACACTGTGTCTCTT	1442
Qy	61	TyrIleuAspSerSerProGluAspAspProThrCysIleThrLysGluPheTyrPalatThr	80
Dp	1443	TACAAAGCTATCAAGTCTCGTAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCACC	1502
Qy	81	IleLeuAspSerAlaGlyProLeuProAspTyrThrProGlnIlePheSerPheGlu	100
Dp	1503	ATTTTGATTCACGAGGTCTCTTCTCGACTATACCTCCCAAAATTTTCTTTTGA	1562
Qy	101	SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly	120



```

Db      1563 AGTACTAGGATTTTACATGTATGGAGTCTTACAGGCTCATGATCTACAGCTTGA 1622
Qy      121 LVSLVSTYrProThrValLeuPheIleTYrGlyGly--ProGln----- 134
Db      1623 AAGAAATATCTTACTGTGTCTTCAATATATGAGTGCTCTCAGGTGAGTGTGTAATA 1682
Qy      134 ----- 134
Db      1683 ATCGGTTTAAAGAGTCAGATATTTCCGCTTGAATACCTAGCCTCTAGATTATGTGG 1742
Qy      134 ----- 134
Db      1743 TTGTAGTATAGACAAGGGAGTCTGTACCGAGGGCTTAATTGGAAGGCCCTTTA 1802
Qy      135 -----GlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTYrLeuAlaS 151
Db      1803 AATATATAATGGGTCAATAATGAATATGACGATCAGGTGGAAGGACTCCAAATATCTAGCTT 1862
Qy      151 eFArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTYrPSeTYrGlyGlyTY 171
Db      1863 CTGGAATATGATTTCACTTGAATGATGCTGTGGGCACTCCAGGCTGGTCTATGAGAGAT 1922
Qy      171 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 191
Db      1923 ACCTCTCCCTGATGAGCATTTATGACAGAGTCAGATATCTTCAGGGTTGCTATTGTGGGG 1982
Qy      191 laProValThrLeuTPrIlePheTYrAspPheTYrThrGlyTYrThrGlyValArgTYrMetGlyHisP 211
Db      1983 CCCCATCTACTCTGTGTGATCTTCTATGATACAGAGATACAGGAACGTTATATGGGTACCC 2042
Qy      211 roAspGlnAsnGlnGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGlnLysPheP 231
Db      2043 CTGACCCAGATGAACAGGGCTATTACTTAGAGATCTGTGGCCATGCAAGCAAGAAAGTCC 2102
Qy      231 roSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheA 251
Db      2103 CCTCTAACCAAAATCGTTTACTCTCTTACATGATGTTCCGTGATGAGAAATGTCATTGTTGG 2162
Qy      251 laHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAspLeuGlnI 271
Db      2163 CACATATACCGATATATTAATCTAGATTTTATGAGAGGCTGGAAGCCCATATGATTTACAGA 2222
Qy      271 leTYrProGlnGlnArgHisSerIleArgValProGlnSerGlyGlnHisTYrGlyLeuH 291
Db      2223 TCTATCCTGAGGAGAGCAACAGCATTAAGATTCCTGAAATCGGAGAAACATTATGAATGCTC 2282
Qy      291 leLeuLeuHisTYrLeuGlnGlnLysAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310
Db      2283 ATCTTTTGACCTACCTTCAGAAACCTTGATCAAGTATGCTGTCTAAAGTGATA 2341

```

```

RESULT 11
AX608735 4829 bp DNA linear PART 17-FEB-2003
DEFINITION Sequence 12 from Patent WO0231134.
ACCESSION AX608735
VERSION AX608735.1 GI:28404303
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
Novel serine protease genes related to dppiv
Patent: WO 0231134-A 12 18-APR-2002;
JOURNAL Ferring BV (NL)
FEATURES
source 1..4829
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1466 a 886 c 1017 g 1460 t

```

## ORIGIN

## Alignment Scores:

Pred. No.:	1,85e-150	Length:	4829
Score:	1625.50	Matches:	310
Percent Similarity:	85.87%	Conservative:	0
Best Local Similarity:	85.87%	Mismatches:	0
Query Match:	96.76%	Indels:	51
DB:	6	Gaps:	1

US-10-070-464-3 (1-310) x AX608735 (1-4829)

```

Qy      1 PheGlnGlyThrLysAspSerProLeuGlnHisLeuTYrValIleSerTYrValAsn 20
Db      1783 TTTGAAGGACCAAGAACTCCCTTATAGAGCATCACCTGAGTACTAGTACGTAAT 1842
Qy      24 ProGlyGlnValIThrArgLeuThrAspArgGlyTYrSerHisSerGlySerGln 40
Db      1843 CCGGAGAGGTGACCAAGGCTGACTGACCGGTGCTACATCTTGTGTCATCAGTCCAG 1902
Qy      41 HisCysAspPhePheIleSerLysTYrSerAsnGlnLysAsnProHisCysValSerLeu 60
Db      1903 CACTGTGACTCTTTTATTAAGTATAGTAAACAGAAAGATCACACTGTGTCTCTT 1962
Qy      61 TYrLysLeuSerSerProGlnAspAspProThrCysLysThrLysGlnPheTrpAlaThr 80
Db      1963 TACAAGCTATCAAGTCTCTGAAGTATGCCCACTTGCAAAACAAAGAAATTTTGGCCACC 2022
Qy      81 lleuAspSerAlaGlyProLeuProAspTYrThrProGlnIlePheSerPheGln 100
Db      2023 ATTTTGATTCAGCAGAGTCTCTCTTCGACTATACCTCCAGAAATTTTCTTTTGAA 2082
Qy      101 SerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGlnProGly 120
Db      2083 AGTACTAGGATTTTCAATTGATGAGGATGCTTACAGACCTATGATCTACAGCTGGA 2142
Qy      121 LysLysTYrProThrValLeuPheIleTYrGlyGly--ProGln----- 134
Db      2143 AAGAAATATCTTACTGTGTCTTCAATATATGAGTGCTCTCCTCAGGTGAGAGTGTGTAATA 2202
Qy      134 ----- 134
Db      2203 ATCGGTTTAAAGAGTCAGATATTTCCGCTGAATACCTAGCCTCTAGATTATGTGG 2262
Qy      134 ----- 134
Db      2263 TTGTAGTATAGACAAGGGAGTCTGTACCGAGGCTTAATTGGAAGGCCCTTTA 2322
Qy      135 -----GlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTYrLeuAlaS 151
Db      2323 AATATATAATGGGTCAATAATGAATAATGACGATCAGGTGGAAGGACTCCAATATCTAGCTT 2382
Qy      151 eFArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTYrPSeTYrGlyGlyTY 171
Db      2383 CTGGAATATGATTTCACTTGAATGATGCTGTGGGCACTCCAGGCTGGTCTATGAGAGAT 2442
Qy      171 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 191
Db      2443 ACCTCTCCCTGATGAGCATTTATGACAGAGTCAGATATCTTCAGGCTGGTCTATGCTGGGG 2502
Qy      191 laProValThrLeuTPrIlePheTYrAspPheTYrThrGlyTYrThrGlyValArgTYrMetGlyHisP 211
Db      2503 CCCCACTACTCTGTGTGATCTTCTATGATACAGATACAGGAACGTTATATATGGGTACCC 2562
Qy      211 roAspGlnAsnGlnGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGlnLysPheP 231
Db      2563 CTGACCCAGATGAACAGGGCTATTACTTAGAGATCTGTGGCCATGCAAGCAAGAAAGTCC 2622
Qy      231 roSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheA 251
Db      2623 CCTCTAACCAAAATCGTTTACTCTCTTACATGATGTTCTCTGATGAGAAATGTCATTGTTGG 2682
Qy      251 laHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAspLeuGlnI 271

```



Db 2683 CACATCCAGTATATAGTATGAGTTTATAGTGAGGCGGCGAAGGACATATGATTACAGA 2742  
 QY 271 TATTTCCTAGAGAGACACAGATAGAGTTCTTAATCGGAGAACATTAAGAACTGC 2802  
 Db 2743 TCTATTCCTAGAGAGACACAGATAGAGTTCTTAATCGGAGAACATTAAGAACTGC 2802  
 QY 291 TATCTTGCACCTACCTTCAAGAAACCTTGATGACGATATGCTGCTTAAAGTATA 2861  
 Db 2803 TATCTTGCACCTACCTTCAAGAAACCTTGATGACGATATGCTGCTTAAAGTATA 2861

RESULT 12  
 BC043124 4799 bp mRNA linear ROD 22-APR-2003  
 LOCUS BC043124  
 DEFINITION Mus musculus dipeptidylpeptidase 8, mRNA (cDNA clone MGC:58057  
 IMAGE:6410075), complete cds.  
 ACCESSION BC043124 GI:27695449  
 VERSION BC043124.1  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 4799)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, R.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carinci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McKean, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richard, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shenvenko, Y.,  
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
 Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE  
 PUBMED 22388257  
 12477932  
 2 (bases 1 to 4799)  
 Strausberg, R.  
 DIRECT SUBMISSION  
 Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Offices, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library Preparation: M. Bento Soares, University of Iowa  
 DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.  
 Thomas L. Casavant.  
 Web site: <http://genome.uiowa.edu>  
 Contact: [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu); [tom-casavant@uiowa.edu](mailto:tom-casavant@uiowa.edu)  
 Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,  
 Flier, K., Keppel, C., Kucada, T., Lebeck, M., Melo, A., Schaefer, K.,  
 Scherz, T., Smith, C., Smith, E., Tack, D., Trout, K., Walters, J.,  
 Casavant, T., Soares, M.B.

REMARK COMMENT  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>  
 Series: Plate: Row: Column: 0.  
 Location/Qualifiers

source 1. 4799  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="MGC:58057 IMAGE:6410075"  
 /issue\_type="Brain, enriched mouse brain 12.5dp"  
 /clone\_lib="NIH BMAP\_F00"  
 /lab\_host="DH10B"  
 /note="Vector: pYX-ASC"  
 1. 4799  
 /gene="Dpp8"  
 /db\_xref="LOCUSID:74388"  
 /db\_xref="MGI:1921638"  
 288 2966  
 /codon\_start=1  
 /product="Dpp8 protein"  
 /protein\_id="AAH43124.1"  
 /db\_xref="GI:27695450"  
 /db\_xref="LOCUSID:74388"  
 /translation="MKIPSGRCNMAAAMFTBOLGVEIFETACEGNGESODRPKLEP  
 FYVEKSMQKLKLLADTRKTHGYMAKAPHPMRYKRTDPDPSDRVLYAMSGEN  
 RENTLPSSEIPKTIKRAVALMSKPLDLFOATLDYGMYSSEELRERKRGTVGI  
 AAYDHPGSGFTLFOAGSGIYHAKDGGPHGFTQOPLRPMVLVETSCNIMDPDLCPAD  
 PMWIAFHNNDIWNILVTRERRLTYVNEELANEDPSAGVATFVLOEEDRYS  
 YWCPQAEPTPSGGLIRLTYEENDESEVEIIVHTSPMLSTRADSRFRPKGTAMPK  
 VPKKQIVLSPVDAAGIIVYDDELVOPELITFEQVYIRACMTPEGKAMSTLDRS  
 QTHLQIVLSPVDAAGIIVYDDELVOPELITFEQVYIRACMTPEGKAMSTLDRS  
 POTHDELLEIFPASCCTGFRILYVITSLKSKYKRSKSGGAPAPDFCPKIEETIT  
 TSGEVAERHGSNINWDEARKLVYFEGKDSPLBHLVYTSANGEVNRLCPAD  
 HSCCSRHDHDFISKSNOKRNVSLYLSPEDDPVHKTKEFMATILDSAGPLDY  
 TPPEIFSEFTSTFTLGMVLYKPHLOPKKTPVLYFGGPOVOLNRRPKGVKFR  
 INTLASIGGVVIVNIRGSCRGKIFEGKFKYKMOIEITDOVEGLOYLASQDFDL  
 DRVGHGNSYGYLISLMALMORSDFRVAIAAPVLMFYDGTGYTERMGHDQEQ  
 GYVLSVAQAQAPFSEPPRLLLHGFLENVHPATSTILSLFLVAKRPDLOIYPO  
 EHSIRVPSGSHYELHLHYLOENIGSRIALKYI"

BASE COUNT 1373 a 1015 c 1103 g 1308 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.74e-148 Length: 4799  
 Score: 1605.50 Matches: 302  
 Percent Similarity: 84.964 Conservative: 3  
 Best Local Similarity: 84.124 Mismatches: 5  
 Query Match: 95.578 Indels: 49  
 DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x BC043124 (1-4799)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyValSerTyValAsn 20  
 Db 1887 TTTGAAGGACCAAGACTCTCTTGGAGCATCAGCTGACGACGATTATGCAAC 1946  
 QY 21 ProGluGluValThrArgLeuThrAspArgGlyTySerHisSerCysCysIleSerGln 40  
 Db 1947 CCTGGAGAGGTGTAGAGCTGACTACCTGGCTACTACACTCCCTGCTCCACCCG 2006  
 QY 41 HisCysAspPhePheIleSerLysTySerAsnGlnLysAsnProHisCysValSerLeu 60  
 Db 2007 CATTGGACTTCTTCAATAGTAAGTACAGCAACAGAAATCAACACTGTGTGCTCC 2066  
 QY 61 TyrLysIleSerSerProGluAspAspProThrCysLysThrLysGluPheThrPalaThr 80  
 Db 2067 TACAAACTTCAGAGCTCTAGAGTACCCAGTTCAATAAACAAGGAATTTGGGCGACC 2126  
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluLysPheSerPheGlu 100  
 Db 2127 ATTGGATTGACGAGGCTCTCTCTGACTACACCCCTCCAGAAATTTTCTTTGAA 2186  
 QY 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyLysPheProHisAspLeuGlnProGly 120  
 Db 2187 ACTACTACTGATTTACACTGTATGTAATGTGTATAGCCCTCATGACCTACCAACCTGGA 2246

```

QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
Db 2247 AAGAAATACCCCACTGTTATTATCATATATGTTGTCCTCCAGGTGACGTTGTAACAAT 2306
QY 134 ----- 134
Db 2307 CCGTTTAAAGAGTCAAGTATTTCCGCTGACACCCCTGCTCCCTGGTTATGTGTT 2366
QY 134 ----- 134
Db 2367 GTGGTATAGACACAGGGGATCTGTACCCGAGACTTAATTGAAAGCGCTTTAA 2426
QY 135 -----GlyGlnIleGlnIleAspAspGlnValGlnIleGlnIleGlnIleAsp 151
Db 2427 TATAAATGGGTCAATATAGAAATCGATGATCAAGTGAAGAGCTCCAGTACTAGCATCT 2486
QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleIleGlyTyrPheTyrGlyGlyTyr 171
Db 2487 CAGTATGACTTCAATGACTTGGATCGAGTGGGATCCACGGCTGCTTATGGTGGCTAC 2546
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
Db 2547 CTCTCCCTATAGGCACTTAATGACAGAGTCCGATATCTCCGGGTGGCTATTTGCGGGCC 2606
QY 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlnArgTyrMetGlyHisPro 211
Db 2607 CCAGTACACCTTGATCTTCTTATGATACAGATATACAGGAGCGCTATATGGCTCACCCCT 2666
QY 212 AspGlnAspGlnGlnIleGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
Db 2667 GACCAAGAAATGAACAGGCTCACTAATGATCTGTGCTCCATGCAAGGCGGAAGTTCCTCC 2726
QY 232 SerGlnProAsnArgLeuLeuLeuLeuLeuIleGlyPheLeuAspGlnValHisPheAla 251
Db 2727 TCGAAGCAAAACGGTTACTCTCTTGCATGGGTCTTGTGATGAAATGTTCACTTTGCA 2786
QY 252 HisThrSerIleLeuLeuSerPheLeuValAlaAlaGlyLysProTyrAspLeuGlnIle 271
Db 2787 CACACAGATATATGCTGATTTTATGATGAGGCTGGAAAGCCATATGATGCTTACAGATC 2846
QY 272 TyrProGlnGlnArgHisSerIleArgValProGlnSerGlyLysIleTyrGlnLeuHis 291
Db 2847 TATCTCAGAGAGGACAGACATCAGAGTCTCTGAGTCTTGGAAACATATGAACTGGAC 2906
QY 292 LeuLeuHisTyrLeuGlnGlnIleAsnLeuGlySerArgIleAlaIleLeuLysValIle 310
Db 2907 CTGCTCACACTTACCTTACAGGAACCTTGATCGGATATCGCTCTGAAAGATGAT 2963
RESULT 13
LOCUS BC030688 3130 bp mRNA linear PRI 30-MAY-2002
DEFINITION Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:26191
IMAGE:4822550, mRNA, complete cds.
ACCESSION BC030688
VERSION BC030688.1 GI:21265132
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3130)
Strausberg, R.
Direct SubMISSION
Submitted (24-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
Toshiyuki and Piero Carninci (RIKEN)

```

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kerteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
0
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAC Plate: 33 Row: d Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18450277.
Location/Qualifiers
1. 3130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26191 IMAGE:4822550"
/issue_type="Test18"
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
372..2667
/codon_start=1
/product="Similar to dipeptidylpeptidase 8"
/protein_id="AAH30688.1"
/db_xref="GI:21265133"
/translation="MAAAMEEOLGVEIFETADCEENIESODRPKLEPFYVERSMO
LKKLADTRKNGYMAKAPHPMPKNDPFGPHDRITRYLAMSENNENTLPSYSEI
PTKINPAALVLMKRPDLDFATLDYGMTSBEELRLRKRKIGTGASIDYHQSQ
TFLFQAGSGIYHVKGQPGFTQOPLRPLVLTSCINIMDPLCPADPMLAFHSH
DIWISNIVTREERRLTYVNNELANMEDRSAGVATFVLOEEDRYSGYWCKAFTT
PGGKILRLYEENDESVEIILHTVSPLMETRADFRYKQGTAPKPTFKSEIM
DAEGRIIDVIDKEILOPELIFEGVYIRACMTBPGKXAMSLILRSOTRLOIVIS
PELRTIPVEDNVERORLISVPSDSTPLIYETTDIWINHIDIFVYPOSHEEIEF
IRASBCKTFPHLKITSLKSKTKSGSGLPASDFPCPIKEEATILSGEVEVGR
HOSNIQVDEVRRLVYEGTKDSPLEHRLVVSIVNGEVTRLDROYSHSCCISQED
PFSIKSNQKPHCVSLYKLSPPEDPTCKTEFMAATILDSAGPLDPYVPEISFES
TTGFTLYGMLYRPHLOPKKYPVLYFVIGGPOVLNNRPFQVKYFRLNTLASLYV
VVVIDRSGCHRGKLEKAPKMYAIAAPVLMIFYDTGYTERYMGHPDQEOGY
LGSVMAOKFSEPRILLILGFLDENHFAHTSLISFLVRAGKPYDLQIYPOGRH
STRVBSGCHYELHLHYIQENDGSRIALAKVI"
BASE COUNT 955 a 626 c 717 g 832 t
ORIGIN
Alignment Scores:
Pred. No.: 4,49e-129 Length: 3130
Score: 1407.00 Matches: 270
Percent Similarity: 88.36% Conservative: 11
Best Local Similarity: 84.91% Mismatches: 19
Query Match: 83.75% Indels: 18
DB: 9 Gaps: 5
US-10-070-464-3 (1-310) x BC030688 (1-3130)
QY 1 PheGlnGlyThrLysAspSerProLeuGlnHisLeuTyrValIleSerTyrValAsn 20
Db 1941 TTTGAAGGACCAAAAGACTCCCTTAAAGCATCACTTATAGTATCGTAAAT 2000
QY 21 ProGlnGlnValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlnSerGln 40
Db 2001 CCTGGAGAGGTGACCAAGGCTGACTGCTGCTCACTTCTTGGCATCACTGAG 2060
QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
Db 2061 CACTGCACTTCTTTTATAGTAAATGATTAACCGAAGAAATCCACACTGTGTCTCTT 2120
QY 61 TyrLysLeuSerSerProGlnAspAspProThrCysLysThrLysGlnPheTrpAlaThr 80
Db 2121 TCAAGGCTATCAAGTCTGAAGATGACCAACTTGGAAAAACAAGAAATTTGGGCCACC 2180
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGlnIlePheSerPheGln 100

```

```

Db      2181 ATTTGGATTACAGACAGTCTCTTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 2240
Qy      101  SerThrThGlyPheThrLeuTyrglyMetLeuTyrlvsProHisAspLeuGlnProgly 120
Db      2241 AGTACTACTGAGATTTCATTGTAAGGAGATGCTTACAGGCTCATATATCTACAGCCTGGA 2300
Qy      121  LysLysTyPProThrValLeuPheIleTyrglyGlyProGlnGlyGlnIleGluIleAsp 140
Db      2301 AAGAAATATCTACTGCTGCTGCTCATATATAGTGCTCTCAGGTGCAGTTG---GTGAAT 2357
Qy      141  AapGlnValGlnGlyLeuGlnIleTyP-----LeuAlaSer---ArgTyraAsp 154
Db      2358 AATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAAATACCTGACCTCTCTAGGTTATGTG 2417
Qy      155  PheIleAspLysAspArgValGlyIle-----HisGlyTyrSerTyrglyTyrlvs 172
Db      2418 GTTGTATGATATACACACAGGGGATCTCTGTACCCAGGGCTTAAATTTGAAGCGCCTTT 2477
Qy      173  SerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPro 192
Db      2478 AATATATAA-----ATGGTTGCTATTGCTGGGGCCCA 2510
Qy      193  ValThrLeuTrpIlePheTyraAspThrGlyTyrlvsGlnArgTyrlvsMetGlyHisProAsp 212
Db      2511 GTCACTCTGTGATCTTCTATGATACAGGATACACGGAACGTTATATGGTCACTCTGAC 2570
Qy      213  GlnAsnGlnGlnGlyTyrlvsTyrlvsGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 232
Db      2571 CAGAAATGAACAGGGCTATTAATTAAGATCTGTGGCCATCGACAGCAAAAGTTCCCTCT 2630
Qy      233  GluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAlaHis 252
Db      2631 GAACCAAAATCGTTTACTCTCTTACATGGTTCTCTGATGAGATGTCATTTTGACAT 2690
Qy      253  ThrSerIleLeuLeuSerPheLeuValAlaGlnGlyLysProTyraAspLeuGlnIleTyP 272
Db      2691 ACCAGTATATTAATCTGATTTTATGAGAGCGCTGGAAAGCCATATGATTTACAGATCTAT 2750
Qy      273  ProGlnGlnArgHisSerIleArgValProGlnSerGlyGlnHisIleTyrglyLeuHisLeu 292
Db      2751 CCTCAGAGAGACACAGCATTAAGAGTTCTTGAAATCCGAGAACATTATGAACGCACTCTT 2810
Qy      293  LeuHisTyrlvsGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 310
Db      2811 TTGCACTACTCTCAAGAAACCTTGATGATCAGGATTTGCTCTTAAATGATGATA 2864

RESULT 14
BD157001      2161 bp      DNA      linear      PAT 17-JAN-2003
LOCUS      BD157001
DEFINITION      Primer for synthesizing full-length cDNA and use thereof.
ACCESSION      BD157001.1 GI:27862759
VERSION      BD157001.1 GI:27862759
KEYWORDS      JP 2002191363-A/11844.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2161)
Otsu,T., Isegai,T., Nishikawa,T., Hayaishi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakematsu,A., Nagai,K., and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 11844 09-JUL-2002;
HELIX RESEARCH INSTITUTE

COMMENT
OS      Homo sapiens (human)
PN      JP 2002191363-A/11844
PD      09-JUL-2002
PF      28-JUL-2000 JP 2000280990
PI      TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI      SAITO,
PI      JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKEMATSU,
PI      KEIICHI NAGAI, TETSUJI OTSUKI
PC      C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC

```

```

10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
PC      Primer for synthesizing full-length cDNA and use thereof FH      key
FEATURES
source      1. 2161
Location/Qualifiers
(16)..(1911).
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT      662 a      429 c      452 g      618 t
ORIGIN
Alignment Scores:
Pred. No.:      1,14e-128      Length:      2161
Score:      1401.00      Matches:      269
Percent Similarity:      88.05%      Conservative:      11
Best Local Similarity:      84.59%      Mismatches:      20
Query Match:      83.39%      Indels:      18
DB:      0      6      Gaps:      5
US-10-070-464-3 (1-310) X BD157001 (1-2161)
Qy      1  PheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyrlvsValSerTyraValAsn 20
Db      988  TTGAAAGGACCAAAAGATCTCCCTTTAGAGCATCACTGATGATGATGATGATGATGAT 1047
Qy      21  ProGlnGlnValThrArgLeuThrAspArgGlyTyrlvsHisSerCysCysIleSerGln 40
Db      1048 CCTGAGAGAGTGAACAAAGCTGACTGACCTGACTACTCATCATCTTCTGTCGATCACTG 1107
Qy      41  HisCysAspPhePheIleSerTyrlvsSerAsnGlnLysAsnProHisCysValSerLeu 60
Db      1108 CACTGTGACTTCTTTATTAAGTAAAGTAAATGATTAACCAAGAAATCCACATCTGTGCTCTT 1167
Qy      61  TyrlvsLeuSerSerProGlnAspAspProThrCysTyrlvsThrLysGlnPheTrpAlaThr 80
Db      1168 TCAAGCTATCAAGCTCTGAATATGATGATGATGATGATGATGATGATGATGATGATG 1227
Qy      81  IleAspSerAlaGlyProLeuProAspTyrlvsProProGlnIlePheSerPheGln 100
Db      1228 ATTTGGATTACAGACAGTCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 1287
Qy      101  SerThrThGlyPheThrLeuTyrglyMetLeuTyrlvsProHisAspLeuGlnProgly 120
Db      1288 AGTACTACTGAGATTTCATTGTAAGGAGATGCTTACAGGCTCATATATGTTGCTCAGGTG 1347
Qy      121  LysLysTyPProThrValLeuPheIleTyrglyGlyProGlnGlyGlnIleGluIleAsp 140
Db      1348 AAGAAATATCTACTGCTGCTGCTCATATATAGTGCTCTCAGGTGCAGTTG---GTGAAT 1404
Qy      141  AapGlnValGlnGlyLeuGlnIleTyP-----LeuAlaSer---ArgTyraAsp 154
Db      1405 AATCGTTTTAAAGGAGTCAAGTATTTCCGCTTGAAATACCTGACCTCTCTAGGTTATGTG 1464
Qy      155  PheIleAspLysAspArgValGlyIle-----HisGlyTyrSerTyrglyTyrlvs 172
Db      1465 GTTGTATGATATACACACAGGGATCTGTCAACGAGGCTTAAATTTGAAGCGCCTTT 1524
Qy      173  SerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPro 192
Db      1525 AATATATAA-----ATGGTTGCTATTGCTGGGGCCCA 1557
Qy      193  ValThrLeuTrpIlePheTyraAspThrGlyTyrlvsGlnArgTyrlvsMetGlyHisProAsp 212
Db      1558 GTCACTCTGTGATCTTCTATGATACAGGATACAGGAACGTTATATGGTCACTCTGAC 1617
Qy      213  GlnAsnGlnGlnGlyTyrlvsTyrlvsGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 232
Db      1618 CAGAAATGAACAGGGCTATTAATTAAGATCTGTGGCCATGACAGCAAAAGTTCCCTCT 1677
Qy      233  GluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAlaHis 252

```

Db 1678 GAACAAATCGTTTACTGCTTACATGCTTCTCGATGAGATGTCATTTTGACAT 1737  
QY 253 ThrSer11LeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyr 272  
Db 1738 ACCAGTATATTCTAGCTTTTATGAGAGGCTGGAAAGCCATATGATTACGATCTAT 1797  
QY 273 ProGlnGluValArgHisSer11LeuArgValProGlnSerGlyGlnHisTyrGluLeuHisLeu 292  
Db 1798 CCTCAGAGAGACACAGCATPAAGATTCCTGATCCGGAGAACATTATGAACTGCATCTT 1857  
QY 293 LeuHisTyrLeuGlnGlnGluLeuGlySerArg11LeuAlaLeuValIle 310  
Db 1858 TTGCCTACCTTCAGAAAACTTGTGATCAGTATGCTGCTCTAAAGTGATA 1911  
RESULT 15  
AK027826 2161 bp mRNA linear PRI 01-ANU-2002  
LOCUS Homo sapiens cDNA FLJ14920 fis, clone PLACE1007416, weakly similar  
DEFINITION to D1PEPTIDYL PEPTIDASE IV (EC 3.4.14.5).  
ACCESSION AK027826  
VERSION AK027826.1 GI:14042789  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Iisogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,  
Wagatsuna,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,  
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,  
Nakamura,Y., Nagahara,K., Masuko,Y., Ninomiya,K. and Iwayanagi,T.  
NEDO human cDNA sequencing project  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 2161)  
REFERENCE 1  
AUTHORS Iisogai,T. and Otsuki,T.  
JOURNAL Direct Submission  
TITLE Submitted (10-MAY-2001) Takao Iisogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology; cDNA library construction,  
5'- & 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.  
FEATURES  
source Location/Qualifiers  
1..2161  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="PLACE1007416"  
/cissue\_type="placenta"  
/clone\_lib="PLACE1"  
/note="Cloning vector: pME18SFL3"  
16..1914  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BA55395.1"  
/db\_xref="GI:14042790"  
/translation="MDPKLCPADPDWIAEIHNSNDIWSNIVTREBRRLTYVHNELANM  
BEDASAGVATEVLOEEDPRYSGYMPCPAETTPSGKILRLIYENDSEVEIHHVT  
SMLETRRADSPRYKPTGTANPKVTFRKMEIMDAGRILIDVIDKELIQPELIFGV  
EYIARAGMTPECKYASILLDRSOTRLOIIVLISPELFIVEDDWMEROLISVPSV  
TFLIYERTDIDWINHDIENHPROSHIEIEIPRSEKTEGRHLYKTSILSKSKY  
KSSGGLPAPSDPCPIKEELIATSGEWLVGHGSNIOVDEVRRLVIEGKDSPLE  
HHLYVSYVNPGEVRLTDRSYSHSCISQHDQDFLSYGNKINPCVSLYKLSPE  
DPCTCKEPMATILDSAGPLDYTPPEIPIFSFTGTGFLYGNLYKPHDLQPEKPYTV  
LFTYGGPOVALNNRKGKGVKPYFLNTALIGVVVVIDRSGCHRLKEGAKPKYAV  
ATAGAVTMTIMIFYDTGYTERWNGHPRQNGGYLGSVANQAKFSPERRLLLHGLFL

DENVFAHTSILSLFLVAKRPDYLIQYPOERHSIRVPESGHEYLHLHYQENIGS  
RIAAALKVI"  
BASE COUNT 662 a 429 c 452 g 618 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,14e-128 Length: 2161  
Score: 1401.00 Matches: 269  
Percent Similarity: 88.05% Conservative: 11  
Best Local Similarity: 84.59% Mismatches: 20  
Query Match: 83.39% Indels: 18  
DB: 9 Gaps: 5  
US-10-070-464-3 (1-310) x AK027826 (1-2161)  
QY 1 PheGlnGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAsp11SerGln 20  
Db 988 TTGAAAGGACCAACAAAGCTCCCTTTAGAGCATCAGCTGTAGTACGTACGTTACGTAAT 1047  
QY 21 ProGlnGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAsp11SerGln 40  
Db 1048 CCTGAGAGAGTGACAAAGGCTGACGTACCGTAGCTACATCTTGTGTCATCAGTCAAG 1107  
QY 41 HisCysAspPhePhe11SerLeuTyrSerAsnGlnLysAsnProHisCysVal1SerLeu 60  
Db 1108 CACTGTGACTCTTTATTAAGTAAGTATGTAACCAAGAAATCCACACTGTGTCTCTT 1167  
QY 61 TyrLeuLeuSerSerProGlnAspAspProThrCysLeuThrLeuGluPheTyrAlaThr 80  
Db 1168 TACAGCTATCAAGTCTCAAGATGACCCACCTGCAACAAAGAAATTTGGGCCACC 1227  
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGlnGlu11PheSerPheGln 100  
Db 122g ATTTGGATTACGACAGGCTCTCTTCTGACTATACCTCCCAAAATTTTCTTTGAA 1287  
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
Db 1288 AGTACTACTGAGATTATCATTTGATGGAGGAGCTCTCAACACCTCATGATCTACAGCTGGA 1347  
QY 121 LysLeuTyrProThrValLeuPheIleTyrGlyGlyProGlnGlu11Leu11LeuAsp 140  
Db 1348 AAGAAATATCTTACTGCTGCTGTCATATGATGGTGGTCTCAGAGTCAGTTG--GTGAAT 1404  
QY 141 AspGlnValGluGlyLeuGlnTyr-----LeuAlaSer---ArgTyrAsp 154  
Db 1405 AATCGGTTTAAAGAGTCAAGATATTTCCGCTGAATACCTTACCCTCTAGATTATGG 1464  
QY 155 Phe11LeuAspLeuAspArgValGlyIle-----HisGlyTyrSerTyrGlyTyrLeu 172  
Db 1465 GTTGTAGTATGACAAACAGGGGATCTGTCAACCGAGGCTTAAATTTGAAGCGGCTTT 1524  
QY 173 SerLeuMetAlaLeuMetGlnArgSerAsp11LeuPheArgValAla11LeuAlaGlyAlaPro 192  
Db 1525 AATATATAA-----ATGGTTCATTCTGGAGGCCCA 1557  
QY 193 Val11LeuTyrP11LeuPheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisAspAsp 212  
Db 1558 GTCACTCTGTGATCTTTATGATACAGGATACACGAAAGTTATATGGGTCAACCTGAC 1617  
QY 213 GlnAsnGluGlnGlyTyrTyrLeuGlnLysSerValAlaMetGlnAlaGlnLysPheProSer 232  
Db 1618 CAGAAATGAACAGGGCTATTACTAGGATCTGGGCCATGCAAGCAAAAGTCCCTCT 1677  
QY 233 GlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnLysValHisPheAlaHis 252  
Db 1678 GAACAAATCGTTTACTGCTCTTACATGCTTCCGAGNAGAAATGTCATTTTGACAT 1737  
QY 253 ThrSer11LeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyr 272  
Db 1738 ACCAGTATATTCTAGCTTTTATGAGAGGCTGGAAAGCCATATGATTTACGATCTAT 1797  
QY 273 ProGlnGluValArgHisSer11LeuArgValProGlnSerGlyGlnHisTyrGluLeuHisLeu 292

Db 1798 CCTCAGGAGAGACACAGCATTAAGAGTTCCTGAATCGGAGACATTATGAACTGCATCTT 1857  
Qy 293 LeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleuIysValIle 310  
Db 1858 TTGCACCTACCTTCAGAGAAACCTTGGATCAGTATTGCTGCTCTTAAAGTGATA 1911

Search completed: October 16, 2003, 00:04:20  
Job time : 3407.49 secs

0

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:14:56 (Search time 227.774 Seconds

(without alignments)  
3673.932 Million cell updates/sec

Title: US-10-070-464-3

Sequence: 1680  
1 FECTKDSPLHLYVSYVN.....HLHLVQENLGRIRALKVI 310

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 510512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus\_p2n.model -DEV=x1h  
-Q=/cgn2.1/USPTO.spool/US10070464/runac\_15102003\_113552\_24799/app\_query.fasta\_1.2652  
-DB=N.Geneseq.15jun03 -QPM=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pcc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10070464 @CCN 1.1 0 @runac\_15102003\_113552\_24799 -NCPUs=6 -ICPU=3  
-NO\_MMAP -LAAGBOUTERY -NEG\_SCORES=0 -WAIT -DSPLBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FCGAPXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.15jun03:\*

1	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1680	100.0	1197	22	AAC85695	Nucleotide sequenc
2	1645.5	97.9	1821	24	ABV76411	Dipeptidyl peptid
3	1645.5	97.9	2643	24	AAH99935	Coding sequence of
4	1645.5	97.9	2649	25	ABX12255	CDNA encoding huma
5	1645.5	97.9	2671	24	ABR83322	CDNA encoding huma
6	1645.5	97.9	3106	24	ABK12892	Human protease PR
7	1645.5	97.9	3120	22	AAC85694	Nucleotide sequenc
8	1645.5	97.9	3120	24	AAD38956	Human dipeptidyl p
9	1645.5	97.9	3143	24	AAH99934	CDNA encoding 195
10	1645.5	97.9	4676	24	ABR83328	CDNA encoding huma
11	1625.5	96.8	4309	24	ABR83328	CDNA encoding huma
12	1625.5	96.8	4829	24	ABR83327	CDNA encoding huma
13	1401	83.4	2151	22	AAH15009	Human CDNA sequenc
14	1377.5	82.0	1669	22	AAC85696	Nucleotide sequenc
15	1377.5	82.0	2668	24	ABN59775	Novel human coding
16	1377.5	82.0	2842	24	ABN59774	Novel human coding
17	1372	81.7	2510	24	AAD23843	Human protease PR
18	1358.5	80.9	4685	24	ABR83332	CDNA encoding huma
19	1352	80.5	4523	24	ABR83325	CDNA encoding huma
20	1094.5	65.1	3287	24	AAD38955	Alternative versio
21	1093.5	65.1	2495	24	AAD38957	Human dipeptidyl p
22	1093.5	65.1	2617	24	ABR83323	CDNA encoding huma
23	1093.5	65.1	2952	24	ABR69090	CDNA encoding huma
24	1093.5	65.1	3024	24	AAD38954	Human dipeptidyl p
25	1093.5	65.1	3716	24	ABG75955	Human PMW encodin
26	1093.5	65.1	4219	24	ABR83335	CDNA encoding huma
27	1093.5	65.1	4302	24	ABR83333	CDNA encoding huma
28	1089.5	64.9	3047	24	ABR69113	CDNA encoding huma
29	1086.5	64.7	1083	22	AAC85697	Nucleotide sequenc
30	1086.5	64.7	4180	24	ABR83339	CDNA encoding huma
31	1086.5	64.7	4263	24	ABR83338	CDNA encoding huma
32	1032.5	61.5	2027	21	AACT7137	Human ORFX ORF262
33	1032.5	61.5	2801	22	AA157896	Human polynucleoti
34	1032.5	61.5	2982	22	AA159666	Human polynucleoti
35	1032.5	61.5	3262	22	AA157880	Human polynucleoti
36	1024.5	61.0	2751	24	AAD38311	Murine dipeptidyl
37	934	55.6	873	22	AAAF8179	Human protease and
38	924	55.0	925	24	ABJ90148	Human polynucleoti
39	872.5	51.9	4076	24	ABR83337	CDNA encoding huma
40	872.5	51.9	4159	24	ABR83336	CDNA encoding huma
41	865.5	51.5	4037	24	ABR83341	CDNA encoding huma
42	865.5	51.5	4120	24	ABR83340	CDNA encoding huma
43	755.5	45.0	823	24	ABR30401	Human G-protein-co
44	708	42.1	3713	23	ABJ10425	Drosophila melanog
45	708	42.1	3783	23	ABJ10641	Drosophila melanog

#### ALIGNMENTS

RESULT 1  
AAC85695  
ID AAC85695 strand; cDNA; 1197 BP.  
XX  
XX AAC85695;  
XX  
XX  
XX 29-UTR-2001 (first entry)  
XX  
XX Nucleotide sequence encoding human DPP8 524Phe-657Gln+707Gly-88211e.  
XX  
XX Human: dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;  
XX dipeptidyl peptidase; DPP1V; T cell; cleavage; diarrhoea;  
XX growth hormone deficiency; glucose level; mucosal regeneration;  
XX non-insulin dependent diabetes mellitus; glucose intolerance;  
XX Immunosuppression; ss.  
XX  
XX Homo sapiens.



PF 12-SEP-2000; 2000CN-0125127.  
 XX  
 PR 12-SEP-2000; 2000CN-0125127.  
 XX  
 PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
 XX  
 PI Mao Y, Xie Y;  
 XX  
 DR WPI; 2002-529809/57.  
 DR P-PSDB; ABB99949.  
 XX  
 PT New dipeptidyl peptidase IV protein 27.5 and encoding polynucleotide,  
 PT useful for treating diabetes and tumors -  
 XX  
 PS Claim 6, Page 24-25 (Disclosure); 32pp; Chinese.



PT New polypeptides 21953, member of human prolyl oligopeptidase family,  
PT useful as diagnostic targets and therapeutic agents for controlling  
PT cancer, lymphoma and leukemia

PS Claim 2, Page-; 121pp; English.

XX This invention relates to an isolated 21953 human prolyl  
CC oligopeptidase, which is cytosolic, antidiabetic, antirheumatic,  
CC antiproteolytic, antithyroid, dermatological, antipsoriatic,  
CC antiasthmatic, ophthalmological, antiinflammatory, nootropic,  
CC antiparinsonian, anticonvulsant, gynecological, vasoregic,  
CC antianemic, cardiant, antihypertensive, anorectic and  
CC metabolic in its action. Uses include gene therapy, expression or  
CC activity of 21953 protein modulator, it is useful for identifying a  
CC compound which binds to it and can be used in preventing, treating  
CC or detecting a cellular proliferative or differentiative disorder.  
CC The 21953 molecules can act as novel diagnostic targets and therapeutic  
CC agents for controlling disorders associated with the aberrant activity  
CC or degradation of peptide hormones e.g., disorders associated with cell  
CC differentiation and proliferation such as cancer, immune function,  
CC reproductive, neurological and cardiovascular function. The 21953  
CC molecules are thus useful for treating and preventing cellular  
CC proliferative and differentiative disorders, haematopoietic neoplastic  
CC disorders, immune disorders such as autoimmune diseases, diabetes  
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,  
CC neuronal disorders, demyelinating diseases, vascular disorders and  
CC metabolism or pain disorders. This sequence represents the cDNA  
CC encoding sequence of 21953 human prolyl oligopeptidase. This  
CC sequence represents the coding sequence of 21953, being the sequence  
CC in between the start and the stop codon of the sequence represented in  
CC AAH99934.

XX SQ Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;

Alignment Scores:

Pid. No.: 8,37e-184 Length: 2643  
Score: 1645.50 Matches: 310  
Percent Similarity: 86.35% Conservative: 0  
Best Local Similarity: 86.35% Mismatches: 0  
Query Match: 97.95% Indels: 49  
DB: 24 Gaps: 1

US-10-070-464-3 (1-310) x AAH99935 (1-2643)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyValValSerTyValAsn 20  
DB 1567 TTGAAAGGACCAAGACTCCCTTAGAGCATCAGCTAGTACGTAAGTAAAT 1626  
QY 21 ProGlyValValThrArgLeuThrAspArgGlyTyrSerHisSerCysValSerGln 40  
DB 1627 CCGAGAGAGGTGACCAAGGCTGACTGACCGGTGCTACTCACTTTGCTGCACTGTCAG 1686  
QY 41 HisCysAspPhePheLeuSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60  
DB 1687 CACTGTGATCTTTTAAAGTAAGTATAGTAAACCAAGAACATCCACACTGTGCTCCTT 1746  
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysValSerThrLysGluPheTyrPalThr 80  
DB 1747 TACAAGCTATCAAGTCTCAAGATGACCAACTTGCAAAAACAAAGAAATTTGGGCCACC 1806  
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100  
DB 1807 ATTGGATTCAAGACAGGCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTTGA 1866  
QY 101 SerThrThrGlyPheThrLeuTyrgLysLeuTyrgLysPheProHisAspLeuGlnProGly 120  
DB 1867 AGTACTACTGGAATTATCATGATGATGGATGCTCTACAAAGCCATGATCTACAGCCTGA 1926  
QY 121 LysLysTyrProThrValLeuPheLeuTyrGlyGlyProGln----- 134  
DB 1927 AAGAAATATCCACTGCTGCTGCTATATGTTGGTCCCTCAGGTGCTGTAATAAT 1986  
QY 134 ----- 134

DB 1987 CCGTTTAAAGAGTCAAGATATTCGCTTGAAATACCCTAGCCTCTAGATTATGTGTT 2046  
QY 134 ----- 134  
DB 2047 GTAGTGATAGACACAGGGGATCTCTGTACCGAGGCTTAAATTTGAGGCCCTTTAAA 2106  
QY 135 -----GlyGlnIleGluIleAspAspGlnValIleGluGlnTyrLeuAlaSer 151  
DB 2107 TATTAATAGGTCATATGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2166  
QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 171  
DB 2167 CGATATGATTTTATGACTATGATGATGATGATGATGATGATGATGATGATGATGAT 2226  
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 191  
DB 2227 CTCTCCCTGATGATGATTAATGAGAGTCAAGATATCTCAAGGTTGCTATTTCTGGGCC 2286  
QY 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
DB 2287 CCACTCAGCTGCTGGATCTTATGATATCAGGATACAGGAACGTTATATGGGTCACT 2346  
QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231  
DB 2347 GACCAAGATGAAACAGGCTATTACTTAGATCTGGCCATGCAAGCAAAAGTTCCCC 2406  
QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAla 251  
DB 2407 TTGAAACCAATTCGTTTACTGCTCTTACATGATTTCTCGATGAGAAATGTCATTTTGA 2466  
QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
DB 2467 CATACAGATATATCTGATGATTTTATGAGAGGCTGGAAACCATATGATTTACAGATC 2526  
QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHis 291  
DB 2529 TATCTCTCGAGAGACACAGCATTAAGATCTCTGATCGGGGAACATTATACATGCAAT 2586  
QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuValIle 310  
DB 2587 CTTTTCACACTCTCAAGAAACCTTGATACGATATTGCTGCTTAAAGTGATTA 2643

RESULT 4  
ABX12255  
ID ABX12255 strand; cDNA; 2649 BP.  
XX  
AC ABX12255;  
XX  
DT 19-MAY-2003 (first entry)  
XX  
DE CDNA encoding human serine protease H1PHUM46.  
XX  
KW Human; ss; gene; H1PHUM46; serine protease; gene therapy; osteoarthritis;  
KW serine protease activity modulation; dipeptidyl peptidase activity;  
KW musculokeletal disease; Hepatitis B virus infection; myotonic dystrophy;  
KW Alzheimer's disease; paraneuronal palsy; Huntington's disease;  
KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;  
KW irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;  
KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;  
KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;  
KW multiple sclerosis; chromosome 19q21-q22.  
XX  
OS Homo sapiens.  
XX  
FH Key 1.2649 location/Qualifiers  
FH CDS  
FT /tag= a  
FT /product= "H1PHUM46"  
FT /note= "Serine protease"  
PN GB2374669-A.  
XX

PD 30-OCT-2002.  
 XX  
 XX 22-JAN-2002; 2002GB-0001404.  
 XX  
 XX 23-JAN-2001; 2001GB-0001760.  
 XX  
 XX (GLAXO) GLAXO GROUP LTD.  
 XX  
 XX Edbrooke MR, Lewis AP;  
 XX WPI: 2003-150703/15.  
 DR P-PSDB; AB007720.  
 XX  
 XX Identifying modulators of serine protease activity useful for treating  
 PT musculoskeletal diseases, by contacting cell expressing a novel serine  
 PT protease polypeptide with a compound and monitoring serine protease  
 PT activity  
 XX  
 XX Claim 12; Page 22-26; 38pp; English.  
 XX  
 XX The invention relates to a method of identifying a substance that  
 CC modulates serine protease activity, comprising contacting a cell such as  
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a  
 CC virus, expressing a serine protease polypeptide (HIPHM 46), or its  
 CC variant having dipeptidyl peptidase activity, or a serine protease  
 CC isolated from the cell with a test substance and monitoring for serine  
 CC protease activity. The method is useful for identifying a substance that  
 CC modulates serine protease activity. A modulator of the serine protease is  
 CC useful in the manufacture of a medicament for treatment or prophylaxis of  
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus  
 CC infection, Alzheimer's disease, paraneuronal nuclear palsy, myotonic  
 CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.  
 CC Additional diseases that may be treated using modulators of the serine  
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung  
 CC disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,  
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,  
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple  
 CC sclerosis. The present sequence represents cDNA encoding the human serine  
 CC protease HIPHM46 which is located on chromosome 15q21-q22.  
 CC  
 SQ Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 8-4e-184 Length: 2649  
 Score: 1645.50 Matches: 310  
 Percent Similarity: 86.35% Conservative: 0  
 Best Local Similarity: 86.35% Mismatches: 0  
 Query Match: 97.95% Indels: 49  
 DB: 25 Gaps: 1  
 US-10-070-464-3 (1-310) x ABX12255 (1-2649)  
 QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuTyValIserTyValAsn 20  
 DB 1570 TTTGAAGGACACCAAGACTCCCTTTAGAGCATCACTGTAGTACGTAAGTAAT 1629  
 QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysValIleSerGln 40  
 DB 1630 CTTGAGAGGTGACCAAGGCTGACTACCGTGGCTACTCACTTCTGCGATCAGTCAG 1689  
 QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValIleSerLeu 60  
 DB 1690 CACTGTGACTTCTTTATAGTAGATAGTACCGAAGAAACCACTGCTGCTCCCTT 1749  
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 80  
 DB 1750 TACAGCATATCAAGGCTCGAAGATGACCACTTCCAAAGAAATTTTGGGCGCAC 1809  
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
 DB 1810 ATTGGATTGACGAGGCTCTTCTCTGACTATACCTCTCCGAAATTTTCTTTTGA 1869  
 QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120

DB 1870 AGTACTACGATTTACATTGATGGATGCTTACCAAGCCCTCATGATTTACAGCTGGA 1929  
 QY 121 LysLysTyrProThrValLeuPheIleTyrGlyProGln----- 134  
 DB 1930 AAGAAATATCCTACTGCTGCTGTTATATATGATGCTCCACAGTGCAGTTGTGAATAT 1989  
 QY 134 ----- 134  
 DB 1990 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTTACGTTATGTGTT 2049  
 QY 134 ----- 134  
 DB 2050 GTAGTAGATAGCAACAGGGAGCTCTGTACCGAGGGCTTAATTTGAAGCCGCTTTAA 2109  
 QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151  
 DB 2110 TATAAATGGGTCAATAGAAATTACATGATCAGTGTGAAGGACCTCAATATCTTACTTCT 2169  
 QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyIleTyr 171  
 DB 2170 CGATATGATTTTCATGACTTGAATGATGTTGGGATCCACGGCTGTCTTATGAGATAC 2229  
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
 DB 2230 CTCTCCCTGATGGCATTTAATGACAGGTCAGATATCTTACAGGTTGCTATTGCTGGGCC 2289  
 QY 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
 DB 2290 CCAGACACTCTGTGGATCTTCTATGATACAGGATACAGAAAGCTTATATGGGTCACTT 2349  
 QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuLysSerValAlaMetGlnAlaGluLysPhePro 231  
 DB 2350 GACCAAGATGAACAGGGCTATTACTTACTAGATCTGTGGCCATGCAAGCAAGAAAGTTCC 2409  
 QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
 DB 2410 TCTGACCAAAATCGTTTACTGCTCTTACATGGTTTCCGTGATGAAATGTCATTGTCGA 2469  
 QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
 DB 2470 CATACAGTATATTTCTGAGTTTATTAGTAGAGGCTGGAAGCCATATGATTTACAGATC 2529  
 QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291  
 DB 2530 TATCTCTGAGAGAGACACAGCATTAAGAGTTCTGTGAATCTGGAGAACATTATGAACTGCAT 2589  
 QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 310  
 DB 2590 CTTTGGCACTTACCTTCAAGAAACCTTGGATCAGTATTTGCTCTTAAAGTGATTA 2646  
 RESULT 5  
 ABRK83322  
 ID ABRK83322 standard; cDNA; 2671 BP.  
 XX  
 AC ABRK83322;  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DPPIV related serine protease DPP-1.  
 XX  
 KW Human, serine protease, dipeptidyl peptidase IV-related protein, DPP;  
 KW DPPIV, infection, human immunodeficiency virus, HIV-1, HIV-2, pain,  
 KW diabetes, infertility, obesity, anorexia, Parkinson's disease, stroke;  
 KW heart failure, hypertension; urinary retention, osteoporosis, cancer;  
 KW ulcer, allergy, cancer, psychotic disorder, neurological disorder;  
 KW dyskinisia; reproductive disorder; inflammatory disorder;  
 KW metabolic disorder; gene; ss.  
 OS Homo sapiens.  
 XX  
 XX W0200231134-A2.

PD 18-APR-2002.  
 XX 12-OCT-2001; 2001WO-USJ1874.  
 XX 12-OCT-2000; 2000US-240117P.  
 PR (FERR ) FERRING BV.  
 PA Qi S, Akinsanya KO, Riviere PJ, Junien J;  
 PI WPI: 2002-444178/47.  
 DR P-PSDB; ABG61591.  
 XX  
 DR New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and  
 PT viral infections, cancers, allergies, neurological disorders, or pain  
 PT  
 Claim 1; Page 53-54; 113pp; English.  
 XX  
 PS The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.  
 XX  
 SQ Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other:  
 Alignment Scores:  
 Pred. No.: 8.5e-184 Length: 2671  
 Score: 1645.50 Matches: 310  
 Percent Similarity: 86.35% Conservative: 0  
 Best Local Similarity: 86.35% Mismatches: 0  
 Query Match: 97.95% Indels: 49  
 DB: 24 Gaps: 1  
 US-10-070-464-3 (1-310) x ABK83322 (1-2671)  
 QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuTyValValSerTyValAsn 20  
 DB 1577 TTTGAAGGACACCAAGACTCCCTTTAGAGCATCACCTGTAGTACGTTACGTTAAAT 1636  
 QY 21 ProGlyIuValThrArgLeuThrAspArgIyTySerHisSerCysCysIleSerGln 40  
 DB 1637 CCTGAGAGGTACAAAGCTGACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1696  
 QY 41 HisCysAspPhePheIleSerLysTySerAsnGlnLysAsnProHisCysValSerLeu 60  
 DB 1697 CACTGTGACTCTTTATATAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1756  
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrLysThr 80  
 DB 1757 TACAAGCTATCAAGCTGCAAGATGACCACTTGCACAAACAAAGAAATTTGGGCCACC 1816  
 QY 81 IleuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGlu 100  
 DB 1817 ATTTTGAATTCAGCAGGTCCTTCTCTGACTATATCTCCGAGAAATTTTCTTTTGA 1876  
 QY 101 SerThrThrGlyPheThrLeuTyrgLysMetLeuTyLysProHisAspLeuGlnProGly 120  
 DB 1877 AGTACTACTGATTTACATTGATGGAGTCTCTCAAGCCCTCATGATCTACAGCCTGGA 1936  
 QY 121 LysLysTyProThrValLeuPheIleTyrgLysGlyProGln----- 134

DB 1937 AAGAAATATCTACTGTCGTCGTTGATATANGTGAGTCTCAGGTGACGTGTAATAT 1996  
 QY 134 ----- 134  
 DB 1997 CGATTTAAAGATGCAAGTATTTCCGCTTGAATACCCAGCTCTTATGTATGTGTT 2056  
 QY 134 ----- 134  
 DB 2057 GTAGTATAGACAAAGGGATCTCTGACCGAGGGCTTAATTTGAAGGCCCTTTAA 2116  
 QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLysLeuIleAsp 151  
 DB 2117 TATAAATGGTCAAAATGAAATGACATGAGTGGAGGACCTCCAAATATCTAGCTTCT 2176  
 QY 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyPheSerTyGlyGlyTy 171  
 DB 2177 CGATATGATTTCAATTTGATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2236  
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
 DB 2237 CTCTCCCTGATGGCATTAAATGACAGAGTCAGATATCTTCAAGGTTGCTATTTGCTGGGCC 2296  
 QY 192 ProValThrLeuThrPheThrLysPheThrGlyTyThrGluArgTyMetGlyHisPro 211  
 DB 2297 CCAATCACTCTGTGATCTTCTATGATACAGAGTACAGGAACGTTATATGGGTCACCT 2356  
 QY 212 AspGlnAsnGluGlnGlyTyThrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231  
 DB 2357 GACCAAGTATGACAGGGCTATTAATCTAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCC 2416  
 QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
 DB 2417 TCTGAAACCAATCGTTTACTGCTCTTACATGCTTCCGATGAGAAATGTCATTTTGA 2476  
 QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAspLeuGlnIle 271  
 DB 2477 CATACCAATATATTAATCTGAGTTTGTAGTGAGGCTGAAAGCCATATGATTTACAGATC 2536  
 QY 272 TyrProGlnIuArgHisSerIleArgValProGluSerGlyGluHisTyrgLysLeuHis 291  
 DB 2537 TATCTCGAGGAGACACACATTAAGACTTCTTAATGGAGAACTATTAAGAACTGAT 2596  
 QY 292 LeuLeuHisTyrgLeuGlnIuAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310  
 DB 2597 CTTTTCGACTACTTCAAGAAACCTTGATCAGTATGCTGCTCTTAAAGTGATA 2653  
 RESULT 6  
 ABK12892  
 ID ABK12892 standard; cDNA; 3106 BP.  
 XX  
 AC ABK12892;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Human protease PRTS-9 cDNA sequence.  
 KM Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;  
 KM cardiovascular; atherosclerosis; autoimmune deficiency syndrome; AIDS;  
 KM inflammatory disorder; acquired immunodeficiency syndrome; AIDS;  
 KM cell proliferative disorder; developmental disorder; epilepsy;  
 KM Duchenne muscular dystrophy; epithelial disorder; neurological disorder;  
 KM reproductive disorder; endometriosis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 203..2851  
 FT /tag= a  
 FT /product= "Human protease PRTS-9"  
 PN MO200198468-A2.  
 XX  
 PD 27-DEC-2001.

```

XX 13-JUN-2001; 2001WO-US19178.
PF 16-JUN-2000; 2000US-212336P.
XX 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM,
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA,
PI Wallis NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Dae D, Kearney L,
PI Kallick DA;
XX
XX WPI; 2002-090437/12.
DR P-PSDB; AAU74749.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 5; Page 166-167; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present nucleic acid sequence encodes
CC the human protease PRTS-9 protein of the invention.
XX
SQ Sequence 3106 BP, 928 A, 633 C, 704 G, 841 T, 0 other;
XX
Alignment Scores:
Pred. No.: 1, 07e-183 Length: 3106
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0
Query Match: 97.95% Indels: 49
DB: 24 Gaps: 1
US-10-070-464-3 (1-310) x ABK12892 (1-3106)
QY 1 PHEGLUGLYTHLYRAPHSEPRLEUGLHLSHLEUTYRVALSERTYRVALAEN 20
DB 1772 TTGGAAGGACCAAGACTCCCTTTAGAGCATCACCTGACTAGTACGTAAGTAAT 1831
QY 21 PROGLVGLVALTHARGLEUTHRPARAGLYTYRSEHISERTYRSEGLIESERGIN 40
DB 1832 CCGAGAGAGGTACCAAGGCTGACTGACCGTGGCTACTCATCTTGTGCTGCACTCAG 1891
QY 41 HISYASPPHEPHEHLESERLYSRSEANGLNLASAPROHISYASVALSERLEU 60
DB 1892 CACTGTGACTTTTATTAAGTATTAATACCAAGAAACCCACACTGTGTCTCTT 1951
QY 61 TYRYSLEUSERSEPRGLUAPAPROTHRCYALYETHLYSLGSLUBHETTPALATHR 80
DB 1952 TACAAGCTATCAAGCTCTGAGATGACCCAACTTGCAAAACAAGAAATTTGGCCACC 2011
QY 81 ILEUAASERVALGLYPROLEUPROASPTYRTHPRPROGLULIEPHESEPRGLU 100
DB 2012 ATTTTGATTCAGCAGGCTCTCTCTGACTACTCTCCACAGAAATTTTCTTTGAA 2071

```

```

QY 101 SetThrThrglyPheThrleuTYRGLYMeLeuTYRlySPROHISAAPLeuGlnProGly 120
DB 2072 AGTACTACTGAGATTTCATTTGATGAGATGCTCTACAGGCTCATGATCTACAGCTCGA 2131
QY 121 LYSLEYTYRPROTHRVALLEUPHEHLEYTYRGLYGLYPROGln----- 134
DB 2132 AAGAAATATCTACTGCTGCTTCATATATATGTTGCTCTCAGGTCGAGTGCATTAAT 2191
QY 134 ----- 134
DB 2192 CGGTTTAAAGAGTCAAGATTTCGCTTGAAATACCTTACCTCTAGTTATGTGTT 2251
QY 134 ----- 134
DB 2252 GTAGTATAGACAAACAGGGATCTCTCACCGAGGCTTAAATTTGAAGCGCTTTAA 2311
QY 135 -----GlyGlnIleGluIleAPAPGlnValGluGlyLeuGlnTYRLeuAlaSer 151
DB 2312 TATTAATGGGTCAATATAGAAATTTGACATCAGTGGAGAGACTCCAAATATCTAGTTCT 2371
QY 192 ARGTYRAPHHEHLESLEUPAPAGVALGLYLIEHISGLYTRPSERTYRGLYGLYTYR 171
DB 2372 CGATATGATTTCAATTGACTTAGATCGTGGGCGATCCAGGCTGCTCTATGAGAGATAC 2431
QY 172 LEUSERLEUMERVALLEUMERGINARGSERAPDILEPHARGVALAlaIleAlaGlyAla 191
DB 2432 CTCTCCCTGATGGCATTTAATGACAGGTCAGATATCTTACGGGTGCTATGCTGGGGCC 2491
QY 192 PROVALTHRLEUTHRILEPHEHLEYRAPHTRHGLYTYRTHRLUARGTYRMEGLYHISPRO 211
DB 2492 CAGGTACTCTGTGATCTTCTATATGATACAGGATACCGGAACGTTATATGGGTCAACCT 2551
QY 212 ASPGINASRGLUGLNGLYTYRILEUGLYSERVALAlaMeGlnAlaGlyLysPhePro 221
DB 2552 GACCAAGAAAGAACAGGGCTATTACTTAGGATCTGTGGCCATGCACAGAAAGATTCCCC 2611
QY 232 SERGLUPROASAPRGLLEULEULEUENHISGLYPHELEUASPGULUENVALHISPEHAla 251
DB 2612 TCTGAACCAAAATGCTTACTGCTCTTACATGGTTTCTCGATGAGAAATGTCATTGTGCA 2671
QY 252 HISRHSERILEULEUSERPHLEUVALARGAlaGlyLysPheProTYRAPHLEUGlnIle 271
DB 2672 CATACAGATATATTACTGAGTTTATAGAGGGCTGGAAGCATATGATTACGATC 2731
QY 272 TYRPROGLINLARGHISERILEARGVALPROGLUSERGLYGLUHISTYRGLULEUHS 291
DB 2732 TATCTCAGAGAGACACAGCATTAAGATTCCGATCCGAGACATTATGAACTGCAT 2791
QY 292 LEULEUHISLYRLEUGLNGLUENLEUGLYSERARGILEAlaAlaLeuLysValIle 310
DB 2792 CTTTTCGACTTCTTCAGAAACCTTGGATCACGATTGCTGCTCTTAAAGTGATA 2848

```

RESULT 7  
 AAC85694  
 ID AAC85694 standard; cDNA; 3120 BP.  
 XX  
 AC AAC85694;  
 XX  
 DT 29-JUN-2001 (first entry)  
 XX  
 DE Nucleotide sequence of human DPP8.  
 XX  
 KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;  
 KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;  
 KW growth hormone deficiency; glucose level; mucosal regeneration;  
 KW non-insulin dependent diabetes mellitus; glucose intolerance;  
 KW immunosuppression; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 214..2862  
 FT /tag= a

/product= "Human DPP8"

MO200119866-A1.

22-MAR-2001.

11-SEP-2000; 2000MO-AU01085.

10-SEP-1999; 99AU-0002762.

18-FEB-2000; 2000AU-0005709.

(UNSY ) UNIV SYDNEY.

Abbott CA, Gorell MD;

WPI: 2001-281520/29.

P-PSDB; AAB87187.

New human dipeptidyl aminopeptidase (DPP8) useful for cleaving substrates, identifying inhibitors of DPP8 catalytic activity which have therapeutic uses, and for detecting activated T cells

Claim 16; Fig 2; 78pp; English.

This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and H-Arg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable of hydrolyzing the peptide bond C-terminal to Pro in each of these compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for cleaving a substrate, and for detecting an activated T cell which involves measuring the level of DPP8 gene expression in a T cell. The level of DPP8 expression is detected by detecting the amount of DPP8 RNA in the cell. It is also useful for identifying a molecule capable of inhibiting the cleavage of the substrate by DPP8. Molecules identified as inhibiting DPP8 catalytic activity may be useful for treating diarrhoea, growth hormone deficiency, lowering glucose levels in non-insulin dependent diabetes mellitus and other disorders involving glucose intolerance, enhancing mucosal regeneration and as immunosuppressants.

Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:

Prod. No.: 1.08e-183 Length: 3120

Score: 1645.50 Matches: 310

Percent Similarity: 86.35% Conservative: 0

Best Local Similarity: 86.35% Mismatches: 0

Query Match: 97.95% Indels: 49

DB: 22 Gaps: 1

US-10-070-464-3 (1-310) x AAC85694 (1-3120)

1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValSerTyrValAsn 20

1783 TTGAAAGGACCAAAAGCTCCCTTTAGACATCACTGACGTAGCTTACGTAAT 1842

21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysSilesGln 40

1843 CCGAGAGAGGTGACAAAGGCTGACGCGGTACTTCACTTCTGTCGACATCACTG 1902

41 HisCysAspPhePheHisSerLysTyrSerAsnGluLysAsnProHisCysValSerLeu 60

1903 CACTGTCGACTCTTTATTAAGTAAAGTAAACCAAGAAATCAACATGTCCTT 1962

61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalaThr 80

1963 TACAGACTTCAAGTCTGTAAGATGACCAACTTGCAAAACAAAGAAATTTTGGCCACC 2022

81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100

2023 ATTTCGATTTCAGACAGTCTCTTCTGACTATATCTCCAGAAATTTTCTTTTGA 2082

0

QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120

DB 2083 AGTACTACTGATTTATTCATTTGATGGATGCTCTTACAAAGCTCATGATCTACAGCCTGGA 2142

QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134

DB 2143 AAGAAATATCTACATGTCGTTCAATATATGATGTCCTCAGGTGAGTGTGAATAAAT 2202

QY 134 ----- 134

DB 2203 CCGTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTAGTTATGTCCT 2262

QY 134 ----- 134

DB 2263 GTAGTAGTACAAACAGGGATCTGTACCCGAGGCTTAAATTTGAGCGCCTTTAAA 2322

QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151

DB 2323 TATAAATAGGTCATATAGAAATTCGATCGATCGATGGAAGAGCTCCAAATATCTAGCTTCT 2382

QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 171

DB 2383 CCGATGATTTTCATTTGACTTCAATGCTGTCGGCATTCACGGCTGTCCTATGAGAGATAC 2442

QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191

DB 2443 CTCTCCCTGATGGCAATTAAATGACAGAGTCAGATATCTTCAGGGTCTATTTGCTGGAGCC 2502

QY 192 ProValThrLeuThrIlePheTyrAspThrArgTyrThrGluArgTyrMetGlyHisPro 211

DB 2503 CCAGTCACCTCTGTGATCTTCTATATATACAGATACAGGAAACGTTATATGGGTACCCCT 2562

QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231

DB 2563 GACCGAAATGAAACAGGCGCTTATCTTAGATCTGTGGCCATGCAAGCAAAAGTTCCCC 2622

QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251

DB 2623 TCTGAACCAAAATCGTTTACTGCTTCACTAGTGTTCCTGAGTAGAATGTCCATTTTGA 2682

QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271

DB 2683 CATACCAATATATATGAGTGTTTTATGAGAGGCTGGAACCATATATATTTACAGATC 2742

QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyLysIleTyrGluLeuHis 291

DB 2743 TATCTCCAGAGACAGACAGCATTAAGAGTCTCTGAATCGGAGAACATTTATGACGTGAT 2802

QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgGlyIleAlaLeuLysValIle 310

DB 2803 CTTTTCGACTACTTCAAGAAACCTTGATCAGCATATGTCGTCTTAAGAGTATA 2859

RESULT 8

AAD38956

ID AAD38956 Standard; cDNA; 3120 BP.

XX

AC AAD38956;

XX

DT 23-SEP-2002 (first entry)

XX

DE Human dipeptidyl peptidase 8 (DPP8) cDNA.

XX

KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis; autoimmunity; human immunodeficiency virus; HIV infection; cytotoxic; graft rejection; antidiabetic; antiinflammatory; immunosuppressive; antiviral; enzyme; gene; ss.

XX

OS Homo sapiens.

XX

Key Location/Qualifiers

FT CDS 214..2862

FT /tag= a

FT 0

/product= "Human DPP8 protein"

```

XX PN MO200234900-A1.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-AU01388.
XX PR 27-OCT-2000; 2000AU-0001078.
XX (UNSY ) UNIV SYDNEY.
XX Abbot CA, Gorrell MD;
XX MPI: 2002-454646/48.
XX P-PSDB; AAE24170.
XX
XX New dipeptidyl peptidase (DPP) peptidases, useful for screening
XX inhibitors of DPP catalytic activity, which may be employed to treat
XX e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV infection -
XX
XX Example: Fig 1; 91pp; English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptidases are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immunodeficiency virus) infection. The present
XX sequence is human DPP8 cDNA.
XX
XX Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1,086-183 Length: 3120
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0
Query Match: 97.95% Indels: 49
DB: 24 Gaps: 1
US-10-070-464-3 (1-310) x AAD38956 (1-3120)
QY 1 PheGluGlyThrThyAspSerProLeuGluHisLeuTyrValSerTyrValaen 20
DB 1783 TTTGAAGGACCAAGACTCCCTTTAGGACATCACTGTAGTGTGCTACGTAAT 1842
QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAllesergin 40
DB 1843 CCTGAGAGGTGACAAAGCTGACTGACCGTGTACTCTCATTTGTGTGATCAGTCA 1902
QY 41 HisCysAspPhePheIleSerTyrSerAsnGluValAspProHisCysValSerLeu 60
DB 1903 CACTGTGACTTTTATATAGTAAGTATATACCCGAAATCCACACTGTGTCCCTT 1962
QY 61 TyrLeuLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 80
DB 1963 TACAAGCATATCAAGCTCTGAATAGCCCACTTGCAAAACAAAGAAATTTGGCCACC 2022
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100
DB 2023 ATTTTGATTTCAGCAGTCTCTCTCTGACTACTCCCTCCAGAAATTTTCTTTTGA 2082
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPheProHisAspLeuGlnProGly 120
DB 2083 AGTACTACTGAGATTATCATGTATGATGAGTCTCTCAAGCTTCATGATCTTACAGCTGGA 2142
QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
DB 2143 AAGAAATATCTACTGTGTGTATATATGTGTGTCTCTCAGGTGACGTGTGTAATAT 2202
QY 134 ----- 134
DB 2203 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATCCCTAGCTCTTAGTATGTGTT 2262

```

```

QY 134 ----- 134
DB 2263 GTAGTATAGACAACAGGGATCTGTGACCGAGGCTTAATTTGAGCGGCTTTAAA 2322
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
DB 2323 TATTAATGGGTCAATATAGAAATTTGACATCAGTGCAGTGAAGACTCCAAATATCTAGCTTCT 2382
QY 162 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 171
DB 2383 CGATATGATTTTCATGACTTAGATCGTGGGACATCCAGCGCTGCTATGAGAGATAC 2442
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB 2443 CTCTCCCGATGAGCTTAATGACAGGTCAGATATCTTGAAGGTGCTATGTGTGGGGCC 2502
QY 192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
DB 2503 CCAATCAGCTGTGATCTTCTATGATACAGATACAGGAACGTTATATGGGTCACTT 2562
QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
DB 2563 GACCAAGATGAACAGGCTATTAAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCTCC 2622
QY 232 SerGluProAsnArgLeuLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
DB 2623 TCTGAACCAATCGTTTACTGTCTTACATGAGTTTCTCGATGAGATATCCATTTTGA 2682
QY 252 HisThrSerIleLeuLeuSerPheLeuValAlaArgAlaGlyLysProTyrAspLeuGlnIle 271
DB 2683 CATACCAATATTTACTGAGTTTATTTAGTAGAGGCTGGAACCATATATATTCAGATAC 2742
QY 272 TyrProGlnIuArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291
DB 2743 TATCTCTCAGAGACACAGCATTAAGTCTCTGAATCGGAGAACATTATGACATGAT 2802
QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLysValIle 310
DB 2803 CTTTTCACCTACCTTCAAGAAACCTGTGATCAGATATTTGCTCTTAAGATGATA 2859
RESULT 9
AAH99934
ID AAH99934 standard, cDNA; 3143 BP.
AC AAH99934;
XX
XX 12-APR-2002 {first entry}
DT
XX
DE cDNA encoding 21953 human prollyl oligopeptidase.
XX
KW 21953 prollyl oligopeptidase; human; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antihypertensive; antidiabetic; antidiabetic;
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 229..2877
XX FT /*tag= a
XX FT /product= "21953 prollyl oligopeptidase"
XX FT /note= "This region is specifically claimed in
XX claim 2"
XX
XX MO200179473-A2.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-US40483.
XX

```

PR 18-APR-2000; 2000US-197508P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Meyers RA, Williamson M;  
 XX  
 DR WPI; 2002-034353/04.  
 XX P-PSDB; AAG78415.  
 PT New polypeptides 21953, member of human prolyl oligopeptidase family,  
 PT useful as diagnostic targets and therapeutic agents for controlling  
 PT cancer, lymphoma and leukemia -  
 XX  
 PS Claim 7; Page 100-102; 121pp; English.  
 CC This invention relates to an isolated 21953 human prolyl  
 CC oligopeptidase, which is cytosolic, antidiabetic, antihypertensive,  
 CC neuroprotective, antihypertensive, dermatological, antiparasitic,  
 CC antidiabetic, ophthalmological, antineoplastic, neurotrophic,  
 CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,  
 CC antianxiety, cardiant, antihypertensive, anorectic and  
 CC metabolic in its action. Uses include gene therapy, expression or  
 CC activity of 21953 protein modulator. It is useful for identifying a  
 CC compound which binds to it and can be used in preventing, treating  
 CC or detecting a cellular proliferative or differentiative disorder.  
 CC The 21953 molecules can act as novel diagnostic targets and therapeutic  
 CC agents for controlling disorders associated with the aberrant activity  
 CC or degradation of peptide hormones e.g., disorders associated with cell  
 CC differentiation and proliferation such as cancer, immune function,  
 CC reproductive, neurological and cardiovascular function. The 21953  
 CC molecules are thus useful for treating and preventing cellular  
 CC proliferative and differentiative disorders, haematopoietic neoplastic  
 CC disorders, immune disorders such as autoimmune diseases, diabetes  
 CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,  
 CC neuronal disorder, demyelinating diseases, vascular disorders and  
 CC metabolism or pain disorders. This sequence represents the cDNA  
 CC encoding sequence of 21953 human prolyl oligopeptidase.  
 XX  
 SQ Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 1 09e-183 Length: 3143  
 Score: 1645.50 Matches: 310  
 Percent Similarity: 86.358 Conservative: 0  
 Best Local Similarity: 86.358 Mismatches: 0  
 Query Match: 97.958 Indels: 49  
 DB: 24 Gaps: 1  
 US-10-070-464-3 (1-310) X AAH99934 (1-3143)

QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134  
 DB 2158 AAGAAATATCTTACTGTGCTGTTATATATGTTGGTCCAGTGTGAGTAATAT 2217  
 QY 134 ----- 134  
 DB 2218 CGGTTTAAAGATCAATATTTCCGCTTGAATACCTAGCCTTAGTTATGTGTT 2277  
 QY 134 ----- 134  
 DB 2278 GTAGTATGACACAGGAGATCTCTGACCGAGGCTTAAATTGAAGCGCTTTAA 2337  
 QY 135 -----GlyGlnIleGluIleAspArgValIleGluIleGluIleTyrLeuAlaSer 151  
 DB 2338 TATTAATATGGGTCAATATGAAATTTACGATCAGGTGAAGAGCTCAATATCTAGCTTCT 2397  
 QY 152 ArgTyrAspPheIleAspLeuAspArgValIleGlyIleHisGlyTyrSerTyrGlyIleTyr 171  
 DB 2398 CGATATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 2457  
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValIleAlaIleAlaGlyAla 191  
 DB 2458 CTCTCCCTGATGCAATTAATGACAGGTCAGATATCTTACGGGTGCTATTTGCTGGAGCC 2517  
 QY 192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
 DB 2518 CCAGTCAGCTGTGATCTTCTATATACAGATACAGGATACAGGATGATATGAGTACCT 2577  
 QY 212 ArgGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231  
 DB 2578 GACCAAGATGAAACAGGCGCTATTAGATCTGTGGCCATCAGACAGAAAGATTCCCC 2637  
 QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
 DB 2638 TTTGAAACCAATCGTTTACTGCTCTTACATGCTTCTGAGATGAGATATTCATTTTGA 2697  
 QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
 DB 2698 CATACCAATATTTACTGATGTTTTTATGAGAGCGCTGAAAGCCATATGATTTACAGATC 2757  
 QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291  
 DB 2758 TATCTCTCAGAGAGACACAGCAATTAAGATTTCTGATCGGAGAACATTTATGAACGAT 2817  
 QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuValIle 310  
 DB 2818 CTTTTCACCTACCTTCAAGAAAACCTTGATACGATTTGCTCTTAAAGTGATA 2874  
 RESULT 10  
 ABRK3331  
 ID ABRK3331 standard; cDNA; 4676 BP.  
 XX  
 AC ABRK3331;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DDRP-1 splice variant #7.  
 XX  
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DDRP;  
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KW dyskinnesia; reproductive disorder; inflammatory disorder;  
 KW metabolic disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200231134-A2.  
 XX  
 PD 18-APR-2002.  
 XX

PF 12-OCT-2001; 2001WO-US31874.  
 XX  
 PR 12-OCT-2000; 2000US-240117P.  
 PA (FERR ) FERRING BV.  
 XX  
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;  
 DR MPI; 2002-444178/47.  
 DR P-PSDB; ABG61500.  
 PR New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
 PT viral infections, cancers, allergies, neurological disorders, or pain  
 XX  
 PS Disclosure; Page 72-73; 11pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related  
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcer, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.  
 XX  
 SQ Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 2,616-183 Length: 4676  
 Score: 1644.50 Matches: 308  
 Percent Similarity: 99.35% Conservative: 0  
 Best Local Similarity: 99.35% Mismatches: 1  
 Query Match: 97.89% Indels: 2  
 DB: Gaps: 1  
 US-10-070-464-3 (1-310) x ABK83331 (1-4676)  
 QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuValValSerTyrValAsn 20  
 DB 1783 TTTGAAGCACCAAGACTCCCTTTAGGACATCACTGTAGTCACTGTTACGTAAT 1842  
 QY 21 ProGlyValValThrArgLeuThrAspArgGlyTyrSerHisSerCysValIleSerGln 40  
 DB 1843 CCTGAGAGGTACAAAGGCTGACCTGACCTGCTACTCATCTTTGCTGCATCGTCA 1902  
 QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60  
 DB 1903 CACTGTGACTCTTTATAGTAAGTAACTACCAAGAAATCCACATGCTGTCCCTT 1962  
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluLuphetrAlaThr 80  
 DB 1963 TACAAAGCTATCAAGCTCGAAGATGACCCCACTTCAAAACAAAGAAATTTTGCCCA 2022  
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
 DB 2023 ATTTTGATTCAGCAGGCTCTTCTTCTGACTATCTCTCCGAAATTTTCTTTTGA 2082  
 QY 101 SerThrThrGlyPheThrLysTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
 DB 2083 AGTACTACTGAGATTACATTTGATAGGATGCTCTCAAGCCTCATGATCTACAGCCTGA 2142  
 QY 121 LysLysTyrProThrValLeuPheLysTyrGlyGlyProGlnGlyGlnIleGluLysLeu 140  
 DB 2143 AAGAAATATCCACTGTCTGTTCATATATGATGCTGCG---GTCAAATAGAAATTTAC 2198

QY 141 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160  
 DB 2199 GATCAGGTGGAGAGACTCCAAATATCTAGCTTTCGATATGATTTCTTACATTAAGTCGT 2258  
 QY 161 ValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180  
 DB 2259 GTGGGCATCCAGGCTGGTCCATATGAGGATACCTCTCCCTGATGSCATTATAGCAGAG 2318  
 QY 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 200  
 DB 2319 TCAGATATCTTCAGAGGTTCCTATGCTGGGCCCCAGTCACTCTGGGATCTTCAATGAT 2378  
 QY 201 ThrGlyTyrThrGluArgTyrMetGlnHisProAspGlnAsnGlnGlnIleTyrTyrLeu 220  
 DB 2379 ACGGATACCGGAAAGCTTATATGGGTCCCTGACCGAAATGAAACAGGGCTATTACTTA 2438  
 QY 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 240  
 DB 2439 GGATCTGTGGCCATCCAAAGCAGAAAGTTCCCTTGAAACCAATCGTTTACTGCTCTTA 2498  
 QY 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 260  
 DB 2499 CATGCTTTCCTGGATGAGAAATGTCATTTTGACATACACAGATATTACTGAGTTTCTTA 2558  
 QY 261 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluAsnHisSerIleArg 280  
 DB 2559 GTGAGGGCTGGAAAGCCATATGATTTACAGATCTATCCACAGAGACACAGCATTAAGA 2618  
 QY 281 ValProGluSerGlyGluHisTyrGluLeuHisLeuHisTyrLeuGlnGluAsnLeu 300  
 DB 2619 GTTCTGATAGGGAGAAACATTATGATGATCTTTTGGCATCTTACTTCAAGAAACCTT 2678  
 QY 301 GlySerArgIleAlaIleLeuLysValIle 310  
 DB 2679 GGATCAGCATTTGCTGCTTAAGTGA 2708  
 RESULT 11  
 ABK83328 0  
 ID ABK83328 standard; cDNA; 4309 BP.  
 XX  
 AC ABK83328;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DPPR-1 splice variant #4.  
 XX  
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;  
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KW dyskinasias; reproductive disorder; inflammatory disorder;  
 KW metabolic disorder; gene; ss.  
 OS Homo sapiens.  
 PN W0200231134-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 12-OCT-2001; 2001WO-US31874.  
 XX  
 PR 12-OCT-2000; 2000US-240117P.  
 XX  
 PA (FERR ) FERRING BV.  
 XX  
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;  
 DR MPI; 2002-444178/47.  
 DR P-PSDB; ABG61597.  
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and



PT viral infections, cancers, allergies, neurological disorders, or pain  
 PT  
 XX  
 PS Disclosure; Page 67-68; 113pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypertension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychophoric and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenic), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABRK3322-ABR83343 encode human DPRP proteins.  
 CC  
 XX  
 SQ Sequence 4309 BP; 1304 A; 790 C; 907 G; 1308 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4 07e-181 Length: 4309  
 Score: 1625.50 Matches: 310  
 Percent Similarity: 85.87% Conservative: 0  
 Best Local Similarity: 85.87% Mismatches: 0  
 Query Match: 96.76% Indels: 51  
 DB: 24 Gaps: 1  
 US-10-070-464-3 (1-310) x ABRK3328 (1-4309)  
 QY 1 PheGluGlyThrLysAspSerProLeuGluHisLysLeuTyrValaSerTyrValaen 20  
 DB 1263 TTGAAGGACCAAGACTCCCTTAGAGCATCACCTGAGTACTAGTACGTAAT 1322  
 QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerGlyLysSerGln 40  
 DB 1323 CCTGAGAGGTGACAAAGCTGACTGACCGTCTACTCATCTTGTGCTGACTGACG 1382  
 QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAspProHisCysValSerLeu 60  
 DB 1383 CACTGTGACTTCTTATATAGTAAAGTATAGTAAACCAAGAAATCCACACTGTGTCTCT 1442  
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGlnPheTyrPalThr 80  
 DB 1443 TACAAGCTATCAAGTCCCTGAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCGACC 1502  
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluLlePheSerPheGlu 100  
 DB 1503 ATTTGGATTTCAGCAGGCTCTCTCTGACTATACCTCCAGAAATTTTCTTTTGAA 1562  
 QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPheHisAspLeuGlnProGly 120  
 DB 1563 AGTACTACTGATTTATACATTGATGAGATGCTTACAAACCTCATGATCTACAGCTGGA 1622  
 QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGln----- 134  
 DB 1623 AAGAAATATCCTACTGTCTGTTCATATATGATGAGTCTCTCAGGTCAGTTGTGAATA 1682  
 QY 134 ----- 134  
 DB 1683 ATCGGTTTAAAGAGTCAAGATATTTCCGCTTGAATACCTTACGCTCTTAGGTATGTGG 1742  
 QY 134 ----- 134  
 DB 1743 TTGTAGTATAGCAACAAGGAGTCTGTACACGAGGCTTAATTTGAAGGCGCTTTA 1802  
 QY 135 -----GlyGlnIleGluLleAspAspGlnValGluGlyLeuGlnTyrLeuAlaAs 151  
 DB 1803 AATATAAATGGGTCAAAATGAATGAGATCAGGTGAGAGACTCCCAATATCTAGACT 1862  
 QY 151 eTAArgTyrAspPheIleAspLeuAspArgValGlyLleHisGlyTyrPseTyrGlyGlyT 171

DB 1863 CTGATATGATTTGACTTGAATCGTGTGGGATCCACGAGCTGCTCTAGAGGAT 1922  
 QY 171 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 191  
 DB 1923 ACCCTCTCCTGATGCGATTATACAGAGGTCAAGATATCTTACAGGTTGCTATGCTGGGG 1982  
 QY 191 IaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisP 211  
 DB 1983 CCCAGTCACTCTGTGATCTTCTATGATACAGATACACGGAACGTTATATGGGTCAACC 2042  
 QY 211 rAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPheP 231  
 DB 2043 CTGACAGAAATACAGGGCTTACTTCAAGATCTGTGGCCATGCAAGAGAAAGCTTCC 2102  
 QY 231 rSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251  
 DB 2103 CTTGACCAAAATGTTTACTGCTCTTACATGAGTTTCTGATGAGATGATGCAATTTGG 2162  
 QY 251 IaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnI 271  
 DB 2163 CACATACAGATATATGATGAGTATTTTAGTGAGGGCTGGAAGCCATATGATTTACAGA 2222  
 QY 271 IeTyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuH 291  
 DB 2223 TCTATCCACAGAGAGACACACATTAAGATTCCGAAACGGAGAACATTATGAACTGC 2282  
 QY 291 IeLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 310  
 DB 2283 ATCTTTGGCACTTACCTTCAAGAAACCTTGATCAGTACGATTTCTGCTTAAGATGATA 2341  
 RESULT 12  
 ID ABRK3327  
 ID ABRK3327 standard; cDNA; 4829 BP.  
 XX  
 AC ABRK3327;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DPRP-1 splice variant #3.  
 XX  
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;  
 KW DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KW metabolic disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200231134-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 12-OCT-2001; 2001WO-US31874.  
 XX  
 PR 12-OCT-2000; 2000US-240117P.  
 XX  
 PA (FERR ) FERRING BV.  
 XX  
 FI Qi S, Akinsanya KO, Riviere PJ, Junien J;  
 XX  
 DR WPI; 2002-44178/47.  
 XX  
 DR P-PSDB; ABR61596.  
 XX  
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
 PT viral infections, cancers, allergies, neurological disorders, or pain  
 PS Disclosure; Page 65-66; 113pp; English.  
 XX

CC The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABK83322-ABK83343 encode human DPPP proteins.

XX Sequence 4829 BP, 1466 A; 886 C; 1017 G; 1460 T; 0 other;

## Alignment Scores:

Score:	4.84e-181	Length:	4829
Percent Similarity:	1625.50	Matches:	310
Best Local Similarity:	85.87%	Conservative:	0
Query Match:	85.87%	Mismatches:	0
	96.76%	Indels:	51
	24	Gaps:	1

US-10-070-464-3 (1-310) x ABK83327 (1-4829)

QY 1 PhegluGlyThrIlyAspSerProLeuGluHisIleuTYrValValSerTYrValAsn 20  
 Db 1783 TTTGAGGACCAACAAAGACTCCCTTTAGACATCCGTGAGTACGTTACCTTAAT 1842  
 QY 21 ProgluGluValThrArgLeuThrAspArgIlyTyrSerHisSerCysCysIleSerGln 40  
 Db 1843 CTTGAGAGGTGACAGGCTGACTGACCGTGGCTTCTACATCTTGTGCTGACGTGAG 1902  
 QY 41 HisCysAspPhePheIleSerIlySerAsnGlnIlyAspProHisCysValSerLeu 60  
 Db 1903 CACTGTGACTCTTTATATAGATAGTAGTAACCAAGAAATCCACACTGTGTCCCTT 1962  
 QY 61 TyrllyleuSerSerProGluAspAspProThrCysIlyThrIlyValPheThrAlaThr 80  
 Db 1963 TACAGACTATCAAGCTCCGGAAGATGACCAACTGCCAACAAGAAATTTTGGCCACC 2022  
 QY 81 IleleuAspSerAlaGlyProleuProAspIlyThrProProGlnIlePheSerPheGlu 100  
 Db 2023 ATTGGAATTCAGCAGGTCCTCTTCTCTACTATCTCTCCAGAAATTTCTCTTTTAA 2082  
 QY 101 SerThrThrGlyPheThrLeuTYrGlyMetLeuTYrIlyAspProHisAspLeuGlnProGly 120  
 Db 2083 AGTACTACTGATTACATTGTATGGATGCTCTACAAAGCCTCATGATCTACAGCTGGA 2142  
 QY 121 LysIlySerProThrValleuPheIleTYrGlyIly--ProGln----- 134  
 Db 2143 AAGAAATATCTTACTGTGCTGTATATATGATGTGTCCTCAGGTGACGTTGTGAATA 2202  
 QY 134 ----- 134  
 Db 2203 ATCGTTAAAGAGTCAAGATTTCCGCTTGAATCCCTTACCTCTAGTTATGTGG 2262  
 QY 134 ----- 134  
 Db 2263 TTGTAGTATAGACAAACAGGGATCTGTGACCGAGGCTTAATTGAAAGCGCCTTTA 2322  
 QY 135 -----GlyGlnIleGluIleAspArgGlnValGluGlyLeuGlnIlyrleuAla 151  
 Db 2323 AATATAAATGGTCAATATGAAATTCAGATCAGTGGAGGACTCCAAATATCTAGCTT 2382  
 QY 151 eArqTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTYrGlyIlyT 171  
 Db 2383 CTCGATATGATTCTTACATTGATCTGTGGGATCCAGCGCTGTCCTTATGAGGAT 2442  
 QY 171 YrleuSerleuMetAlaIleuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 191

Db 2443 ACCTCTCCGTGATGATTAATGACAGGTGACAGATATCTTCAGGGTGTCTATTGCTGGG 2502  
 QY 191 laProvalThrleuThrIlePheTYrAspThrGlyTYrThrGluArgTYrMetGlyHisP 211  
 Db 2503 CCCAGTACCTGTGGATCTTCTATGATGATACAGATACCGAAGCTTATATGAGGTACCC 2562  
 QY 211 roAspGluAspGlnGlnGlyTYrTYrleuGlySerValAlaMetGlnAlaGluIlyPheP 231  
 Db 2563 CTGACCAATATGACAGGCTATTACTTACGATCTGTGCTGATGCAATGCAAGAAAGTTCC 2622  
 QY 231 roSerGluProAspArgleuLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251  
 Db 2623 COTGGAACCAATGCTTATCTGCTTACATGATGTTCTGTGATGAGATGTCATTGTTG 2682  
 QY 251 laHisThrSerIleleuLeuSerPheLeuValArgAlaGlyIlyAspProTYrAspLeuGln 271  
 Db 2683 CACATACCAATATATATCTGAGTATTTTATGAGGCTGGAAGCAATATGATTTTACAGA 2742  
 QY 271 leTYrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTYrGlyLeuH 291  
 Db 2743 TCTATCTCAGAGAGACACAGCATTAAGATTCTGAAATCCGGAACATTATGAACTGC 2802  
 QY 291 lsleuLeuHisTYrleuGlnGluAsnLeuGlySerArgIleAlaAlaIleuIlyValIle 310  
 Db 2803 ATCTTTGACATCTACCTTCAAGAAACCTTGGATGACGTAATGCTGCTCTAAAAAGTATA 2861

RESURF 13 0 4  
 AAH15009 standard; cDNA; 2161 BP.  
 XX ID AAH15009;  
 AC AAH15009;  
 XX 26-JUN-2001 (first entry)  
 DT 26-JUN-2001 (first entry)  
 XX Human CDNA sequence SEQ ID NO:12963.  
 DE Human;  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 OS Homo sapiens.  
 PN EPI074617-A2.  
 XX 07-FEB-2001.  
 PD 28-JUL-2000; 2000EP-0116126.  
 PF 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0189767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 PA (HELI-) HELIX RES INST.  
 PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Iishi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 PT primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX Claim 8; SEQ ID 12963; 2537bp + CD ROM; English.  
 PS The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification, where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

SO Sequence 2161 BP; 662 A; 429 C; 452 G; 618 T; 0 other;

# Alignment Scores:

Pred. No.:	Length:	2161
Score:	1401.00	269
Percent Similarity:	88.05%	11
Best Local Similarity:	84.59%	20
Query Match:	83.39%	18
DB:	22	5

US-10-070-464-3 (1-310) x AAH15009 (1-2161)

OY 1 PheGluGluThyThyAAspSerProLeuGluHisLeuTyrValValSerTyrValaen 20  
 DB 988 TTGAAAGGACCAAAAGACTCCCTTTAGAGCATCAGCTTACGTAAGTACGTTAAGTAAAT 1047  
 OY 21 ProGluGluValThrArgLeuThrAAspArgGlyTyrSerHisSerCysCysIleSerGln 40  
 DB 1048 CCTGGAGAGGTGACAAAGCTGACGACGAGTACTGACATTCATCTTGTCGATCAGTACG 1107  
 OY 41 HisCysAAspPhePheIleSerTyrSerAAsnGlnLysAAspProHisCysValSerLeu 60  
 DB 1108 CACTGAGACTCTTTATTAAGTAAAGTAAATCAACAAGAACCAACACGTCGTGCTCCCT 1167  
 OY 61 TyrLysLeuSerSerProGluAAspProThrCysValSerThrLeuGluPheTrpAlaThr 80  
 DB 1168 TACAAAGCTTCAAGTCTGTAAGATGACCCAACTTGCAAAACAAGAAATTTGGGCGCAC 1227  
 OY 81 IleLeuAAspSerAlaGlyProLeuProAAspTyrThrProGluIlePheSerPheGlu 100  
 DB 1228 ATTTGAGATTGACAGAGTCTCTCTCTGACTACTACTCCCAAAATTTCTCTTTTGA 1287  
 OY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAAspLeuGlnProGly 120  
 DB 1288 AGTACTACTGATTTTACATTTGATGAGATGCTTACAAAGCTCATGATCTTACAGCCTGGA 1347  
 OY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnGlyGlnIleGluIleAAsp 140  
 DB 1348 AAGAAATATCCTACTGCTGCTGTTATATGTTGTCCTCAGTGCAGTTG---GTGAAT 1404  
 OY 141 AspGlnValGluGlyLeuGlnTyr-----LeuAlaSer---ArgTyrAAsp 154  
 DB 1405 AATCGGTTAAAGAGTCAAGTATTTCCGCTTGAATACCTCCTAGCTCTAGATTATGAG 1464  
 OY 155 PheIleAAspLeuAAspArgValGlyIle-----HisGlyTyrSerTyrGlyGlyTyrLeu 172  
 DB 1465 GTTGATGATGATGACAAAGGAGATCTCTGACCGAGGCTTAAATTTAAAGCGGCTTT 1524  
 OY 173 SerLeuMetAlaLeuMetGlnAAspSerAAspIlePheArgValAlaIleAlaGlyAlaPro 192  
 DB 1525 AAATATAA-----ATGTTGTCATATTCCTGGGGCCCA 1557  
 OY 193 ValThrLeuThrIlePheTyrAAspThrGlyTyrThrGluAAspTyrMetGlyHisProAAsp 212  
 DB 1558 GTCACCTGCTGATCTCTATATATACAGATACACGGAACGTTAATATGGGTACACCTGAC 1617  
 OY 213 GlnAAsnGluGlnGlyTyrTyrLeuGlySerValAlaIleMetGlnAlaGluLysPheProSer 232

DB 1618 CGAATGAAACAGGGCTATTACTTAGATCTGTGGCCATGCAAGCAAGAAAGTCCCTCT 1677  
 OY 239 GIUPTDAAsnArgLeuLeuLeuLeuHisGlyPheLeuAAspGluAAsnValHisPheAlaHis 252  
 DB 1678 GAACCAATCGTTTACTGCTCTTACATGGTTTCTCGATGATGAAGATCCATTGTCACAT 1737  
 OY 253 ThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAAspLeuGlnIleTyr 272  
 DB 1738 ACCAGATATTACTGATGTTTATTAGAGGCGCTGGAAGCCATATGATTACAGATCTAT 1797  
 OY 273 ProGlnGluAAsnHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHisLeu 292  
 DB 1798 CCTCAGAGAGACACAGCATTAAGATCTCTGATTCGGGGAACATTATGAACTGCATCTT 1857  
 OY 293 LeuHisTyrLeuGlnGluAAsnLeuGlySerArgGlyLysAlaIleAlaLeuLysValIle 310  
 DB 1858 TTGCACTACTCTTCAAGAAACCTTGATCAGTATTTGCTGCTTAAAGTGATA 1911

# RESULT 14

AAAC85696  
 ID AAC85696 standard; cDNA; 1669 BP.

XX AAC85696;

XX 29-JUN-2001 (first entry)

XX Nucleotide sequence encoding human DPP8 318Thr-658Val+759Ala-882Ile.

XX Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;

XX dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;

XX growth hormone deficiency; glucose level; mucosal regeneration;

XX non-insulin dependent diabetes mellitus; glucose intolerance;

XX immunosuppression; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 2..1399

XX FT /tag= a

XX PN /product= "Human DPP8 318Thr-658Val+759Ala-882Ile"

XX MO200119866-A1.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000MO-AU01085.

XX 10-SEP-1999; 99AU-0002762.

XX 18-FEB-2000; 2000AU-0005709.

XX (UNSY ) UNIV SYDNEY.

XX Abbot CA, Gorell MD;

XX WPI; 2001-281520/29.

XX P-PSDB; AAB47189.

XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving

XX PT substrates, identifying inhibitors of DPP8 catalytic activity which

XX have therapeutic uses, and for detecting activated T cells -

XX Claim 21; Page 72-74; 78pp; English.

XX The sequences given in AAC85696-97 encode fragments of human

XX dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for

XX H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a

XX prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable

XX of hydrolysing the peptide bond C-terminal to pro in each of these

XX compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for

XX cleaving a substrate, and for detecting an activated T cell which

XX CC involves measuring the level of DPP8 gene expression in a T cell. The

XX level of DPP8 expression is detected by detecting the amount of DPP8

CC RNA in the cell. It is also useful for identifying a molecule capable  
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules  
 CC identified as inhibiting DPP8 catalytic activity may be useful for  
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels  
 CC in non-insulin dependent diabetes mellitus and other disorders  
 CC involving glucose intolerance, enhancing mucosal regeneration and  
 CC as immunosuppressants.

XX Sequence 1669 BP; 524 A; 336 C; 336 G; 473 T; 0 other;

#### Alignment Scores:

Pred. No.:	2,09e-152	Length:	1669
Score:	1377.50	Matches:	259
Best Similarity:	83.55%	Conservative:	0
Best Local Similarity:	83.55%	Mismatches:	0
Query Match:	81.99%	Indels:	51
DB:	22	Gaps:	1

US-10-070-464-3 (1-310) x AAC85696 (1-1669)

```

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValValSerTyrValaAn 20
DB 620 TTGAAAGGACCAAGACTCCCTTTAGACATCAGCTGAGTACGTTACGTAAT 679
QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGin 40
DB 680 CCTGGAGAGGTACAAAGCTGACTGACCTGGCTACTCATCTTCTGGCATCAGTACG 739
QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
DB 740 CACTGTGACTTTTATTAAGTAAGTAACTAACCAAGAAATCCACACGTGTGTCCTT 799
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysValSerThrLysGlnPheThrAlaThr 80
DB 800 TACAAAGCTATCAAGTCTTAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCCACC 859
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100
DB 860 ATTTGGATTGACGAGGTCTCTCTGCTACTATCTCCCAAAATTTCTCTTTGAA 919
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
DB 920 AGTACTACTGAGATTATCATTTGATGGATGCTCTACAAACCTCATGATCTTACAGCTCGA 979
QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnGlyIleGlnIleAsp 140
DB 980 AAGAAATATCTACTGCTGCTGTTCAATATGTTGGTCTCTCAG----- 1021
QY 141 AspGlnValGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160
DB 1021 ----- 1021
QY 161 ValGlyIleHisGlyTyrPserTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180
DB 1021 ----- 1021
QY 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAsp 200
DB 1022 -----CTTCTTATGCTGGGGCCCAAGTACTCTGTGATCTTCTATGAT 1066
QY 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeu 220
DB 1067 ACAGATACACGGAAGCTTATGATGAGTCACTGACCAAGATGAACAGAGGCTATTACTTA 1126
QY 221 GlySerValAlaMetGlnAlaGluLysPheProSerGlnProAsnArgLeuLeuLeu 240
DB 1127 GGATCTGGCCATGCAAGACAAAGATTCCTCTGAAACCAATCGTTTACTGCTCTTA 1186
QY 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 260
DB 1187 CATGCTTTCCTGGAGAGAAATGTCATTGTGCACATACAGATATTAATGAGTTTITTA 1246
QY 261 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 280

```

```

DB 1247 GTGAGGGCTGGAAGACCATATGATTATCAGATCTATCTCAGAGACACACATTAAGA 1306
QY 281 ValProGluSerGlyGlnHisTyrGluLeuHisLeuHisTyrLeuGlnGlnAsnLeu 300
DB 1307 GTTCTGGAATCGGAGAAACATTAATGAACTGATCTTTTGCACATCACTTCAAGAAACCTT 1366
QY 301 GlySerArgIleAlaAlaLeuLysValIle 310
DB 1367 GCATCAAGTATGCTGCTCTTAAGATGATTA 1396

```

#### RESULT 15

ABNS9775  
 ID ABNS9775 standard; cDNA; 2668 BP.

XX ABNS9775;

XX 28-JUN-2002 (first entry)

XX Novel human coding sequence SEQ ID NO: 186.

XX Human; anti-nausea; vulnary; anti-inflammatory; immunomodulator;

XX anti-fertility; cerebroprotective; cytoskeletal; rheumatic; gene therapy;

XX neuroprotective; antiparkinsonian; protein therapy; EST;

XX expressed sequence tag; gene; ss.

XX Homo sapiens.

XX WO200222660-A2.

XX 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US26015.

XX 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX Xue AJ, Yang Y, Weinman T, Drmanac RT;

XX WPI; 2002-292408/33.

XX P-PSDB; ABB97362.

XX An isolated polynucleotide for treating diseases associated with its

XX encoded polypeptide such as cancer and multiple sclerosis -

XX Claim 1; SEQ ID NO 186; 509bp; English.

XX The present invention provides the protein and coding sequences of 444

XX novel human proteins. These were isolated from expressed sequences tags

XX (ESTs). They can be used to stimulate cell growth, to help tissue regrowth

XX e.g. in burn treatment, to regulate the immune system e.g. to treat

XX multiple sclerosis, to regulate activin or inhibin e.g. to treat

XX infertility, to regulate haemostasis or thrombolysis e.g. to treat

XX stroke and cancer, to screen for drugs, to treat inflammatory conditions

XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

XX Parkinson's disease. The present sequence is a coding sequence of the

XX Sequence 2668 BP; 796 A; 564 C; 592 G; 716 T; 0 other;

#### Alignment Scores:

Pred. No.:	4,26e-152	Length:	2668
Score:	1377.50 <td>Matches:</td> <td>259 </td>	Matches:	259
Best Similarity:	83.55% <td>Conservative:</td> <td>0 </td>	Conservative:	0
Best Local Similarity:	83.55% <td>Mismatches:</td> <td>0 </td>	Mismatches:	0
Query Match:	81.99% <td>Indels:</td> <td>51 </td>	Indels:	51
DB:	24	Gaps:	1

US-10-070-464-3 (1-310) x ABNS9775 (1-2668)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValIserTyrValaen 20  
DB 1629 TTTGAAGGACCAAAAGACTCCCTTAGAGCATCACCTGACGTAGTACGTAAAT 1688  
QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysValIserGln 40  
DB 1689 CCGGAGAGGTGCAAGGCTGACTGACCGGTGCTACTCAATTCCTGTCATCACTCAG 1748  
QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60  
DB 1749 CACTGTGACTTCTTTATAGTAGTATAGTAAACCAAGATCCACACTGTGTCTCTT 1808  
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 80  
DB 1809 TACAAGCTATCAAGTCTCAAGATGACCAAACTTGCAAAACAAAGAAATTTGGCCACC 1868  
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
DB 1869 ATTTGGATTCCAGCAGGCTCTCTTCTGACTATCTCTCCAGAAATTTCTCTTTGAA 1928  
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
DB 1929 AGTACTACTGAGATTTCATTTGATGGATGCTCTACAGCCTCATGATCTACAGCCTGA 1988  
QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnGlyGlnIleGluIleAsp 140  
DB 1989 AAGAATATATCCATCTGTGCTGATATATGGTGTCTCAG----- 2030  
QY 141 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160  
DB 2030 ----- 2030  
QY 161 ValGlyIleHisGlyTyrPserTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180  
DB 2030 ----- 2030  
QY 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 200  
DB 2031 -----GTTGCTATGCTGGGCCCCCAGTCACCTGTGATCTTCTATGAT 2075  
QY 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeu 220  
DB 2076 ACRGATACACGGAAGCTTATATGGGTCAACCTGACCAAGATGAACAGGGCTATTTACTTA 2135  
QY 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 240  
DB 2136 GGATCTGTGGCCATGCAAGCAGAAAGTCCCTCTGAACCAAAATGTTACTGCTCTTA 2195  
QY 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 260  
DB 2196 CATGGTTTCTCGATGAGATGTCATTTTGACACATCACATACCATATTTACTGAGTTTCTTA 2255  
QY 261 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 280  
DB 2256 GTGAGGGCTGGAAGCATATGATTTACAGATCTATCTCAGGAGAGACACAGCATTAAGA 2315  
QY 281 ValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 300  
DB 2316 GTTCCGAAATCGGAGAAACATTAATGAACTGCACTTTTGCCTACTACTTCAGAAAAACCTT 2375  
QY 301 GlySerArgIleAlaIleLeuLysValIle 310  
DB 2376 GGATCAGTATTTGCTCTCTAAAGTGATA 2405

Search completed: October 15, 2003, 17:57:12

Job time : 249.774 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:29:02 ; Search time 48.5672 Seconds  
(without alignments)  
2817.309 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1680

Sequence: 1 FEGTKDSDPLEHLYVVSYYN.....HLHYIQENLGRIMALKVI 310

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-O=/cgn2\_1/USPTO.spool/US10070464/rn1at 15102003 113554 24845/app query.fasta\_1.2652  
-DEL=frame+ p2n.model -DEV=x1h  
-DB=Issued\_Patents\_NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi  
-LIST=45 -DOCLIN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10070464 @CGN 1.1 146 @runat 15102003 113554 24845 -NCPU=6 -ICPU=3  
-NO MMP -LARGEJQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	755.5	45.0	823	4	US-09-280-116-171 Sequence 171, App
2	405.5	24.1	612	4	US-09-392-184-31 Sequence 31, Appl
3	337.5	20.1	3407	4	US-10-002-593-5 Sequence 5, Appl
4	333.5	19.9	2924	5	PCT-US93-07923-1 Sequence 1, Appl
5	337	19.5	5486	4	US-09-462-284-1 Sequence 1, Appl
6	302	18.0	2815	1	US-08-220-491A-1 Sequence 1, Appl
7	302	18.0	2815	1	US-08-619-280A-1 Sequence 1, Appl
8	302	18.0	2815	2	US-08-940-391-1 Sequence 1, Appl
9	267	15.9	4982	3	US-08-699-103B-1 Sequence 1, Appl
10	267	15.9	4982	4	US-09-229-059-1 Sequence 1, Appl
11	267	15.9	4982	4	US-09-628-133-1 Sequence 1, Appl
12	263	15.7	543	4	US-09-221-017B-253 Sequence 253, App

C	13	252	15.0	657	4	US-09-221-017B-646	Sequence 646, App
	14	210	12.5	535	4	US-09-389-681-428	Sequence 428, App
	15	210	12.5	535	4	US-09-620-405B-428	Sequence 428, App
	16	210	12.5	535	4	US-09-433-826B-428	Sequence 428, App
	17	210	12.5	535	4	US-09-604-287A-428	Sequence 428, App
	18	207	12.3	1896	3	US-09-016-080-2	Sequence 2, Appl
	19	193	11.5	502	4	US-09-280-116-172	Sequence 12, App
C	20	159.5	9.5	3085	4	US-09-221-017B-1010	Sequence 1010, App
	21	143.5	8.5	1869	2	US-08-664-646A-1	Sequence 1, Appl
	22	143.5	8.5	1869	2	US-09-066-285-1	Sequence 1, Appl
	23	143.5	8.5	1869	3	US-09-261-006-1	Sequence 1, Appl
	24	143.5	8.5	1869	3	US-08-951-088-1	Sequence 1, Appl
	25	143.5	8.5	1869	4	US-09-609-566-1	Sequence 1, Appl
	26	143.5	8.5	1869	4	US-09-609-570-1	Sequence 1, Appl
	27	143.5	8.5	1869	4	US-09-427-372-1	Sequence 1, Appl
	28	143.5	8.5	1869	4	US-09-693-554-1	Sequence 1, Appl
	29	138	8.2	1845	4	US-09-252-991A-3489	Sequence 3489, App
C	30	110.5	6.6	1230025	4	US-09-198-452A-1	Sequence 1, Appl
	31	110	6.5	904	4	US-09-221-017B-58	Sequence 58, App
	32	96.5	5.7	815	4	US-09-221-017B-962	Sequence 962, App
	33	95.5	5.7	1032	4	US-09-724-623-17	Sequence 17, Appl
	34	95	5.7	679	4	US-09-221-017B-13	Sequence 13, Appl
	35	94	5.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl
	36	94	5.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
	37	92.5	5.5	12835	4	US-08-961-527-141	Sequence 141, App
	38	91.5	5.4	1710	3	US-09-668-960-14	Sequence 14, App
	39	91	5.4	3601	4	US-09-668-960-12	Sequence 2, Appl
	40	90.5	5.4	1389	4	US-09-134-001C-1463	Sequence 1463, App
C	41	90	5.4	35081	2	US-08-752-760A-1	Sequence 1, Appl
	42	89.5	5.3	1830121	4	US-09-557-884-1	Sequence 1, Appl
	43	89.5	5.3	1830121	4	US-09-643-990A-1	Sequence 1, Appl
	44	89	5.3	2046	1	US-07-688-299-12	Sequence 12, Appl
	45	89	5.3	2046	1	US-07-980-517A-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-09-280-116-171  
; Sequence 171, Application US/09280116A  
; Patent No. 6331427  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs  
; FILE REFERENCE: 5800-24, 035800/176965  
; CURRENT APPLICATION NUMBER: US/09/280, 116A  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 268  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 171  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Proyl1 oligopeptidases  
US-09-280-116-171

#### Alignment Scores:

Pred. No.: 6.56e-80  
Score: 755.50  
Percent Similarity: 88.95%  
Best Local Similarity: 86.32%  
Query Match: 44.97%  
DB: 4  
Length: 823  
Matches: 164  
Conservative: 5  
Mismatch: 5  
Indels: 16  
Gaps: 2

US-10-070-464-3 (1-310) x US-09-280-116-171 (1-823)

QY 136 GmltGluilEaspApGlnValGlugly-LeuGlnTYrLeuAlaSerArgTYrAspp 155  
Db 3 CAATGGAATTCACCATCAGTGGAGAGACATCCATATCTTACGCTTCGATATATTT 62  
QY 155 eLlEaSpLpApGlnValGlylLehIsGlyTYrSerTYrGlyGlyTYrLeuSerLeu 175  
|||||

```
Db 63 CATTGACCTAGATGCTGTGGGATCCAGCGCTGGTCTATGAGAGATACCTCTCCCTGAT 122
Qy 175 TAlaLeuMeGlnArGserAspIlePheArGValAlaIleAlaGlyAlaProValThrIle 195
Db 123 GGCATTAAAGCAGAGGTCAGATATCTTCAAGGGTTCCTTCTGGGTCCTCCAGTCACTCT 182
Qy 195 UTPIlePheTyArPThrGlyTyThrGluArGTyMeGlyYHISProAspGlnAsnGly 215
Db 183 GTGATCTTCTATATACAGAGATACAGGAACGTTATATGGGTCACTCCCTGACCAAGATGA 242
Qy 215 UGlnGlyTyTyTyLeuGlySerValAlaMeGlnAlaGlyLysPheProSerGluProAs 235
Db 243 ACAGAGGCTATTAAGTATGATCTGTGGCCATCCAGCAAGAAAGTTCCCTCTGAAACCAA 302
Qy 235 nArGLeuLeuLeuYHISGlyPheLeuAspGluAsnValHISpHeAlaHISThrSerIle 255
Db 303 TCGTTACTGCTCTTACATGATGTTCTCGATGAGAAATGTCATTTCGACATACAGAT 362
Qy 255 eLeuLeuSerPheLeuValArGAlaGlyLysProTyArSpleuGln--IleTy----- 272
Db 363 ATTAAGTATGTTTAAAGAGGCTGGAACAGATATATTACAGTATCTTATATTTT 422
Qy 273 -----ProGlnGluArGHis-SerIleArGVal-ProGluSer-Gl 285
Db 423 GTTGTGTGTAAGATCTATCTCAGAGGAGAGACCCGCGATTAAGATTCCTGAATCGGG 482
Qy 285 YGluHISTyGlyLeuYHISLeuLeu-HISTyTyLeuGlnGlu-AsnLeuGlySer-ArG1 304
Db 483 AGAACATTATTAACATGATCTTTTGGCAGTACCTTCAAGAAACCTTGGATCACTTAT 542
Qy 304 eAlaAlaLeuLysVal 309
Db 543 TGCTGCTCTTAATAATT 558

RESULT 2
US-09-392-184-31/C
; Sequence 31, Application US/09392184
; Patent No. 6395889
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392,184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: prolyl oligopeptidase)
; NAME/KEY: misc feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-392-184-31

Alignment Scores:
Pred. No.: 1.02e-38 Length: 612
Score: 405.50 Matches: 100
Percent Similarity: 67.07% Conservative: 10
Best Local Similarity: 60.98% Mismatches: 49
Query Match: 24.14% Indels: 7
DB: 4 Gaps: 1

US-10-070-464-3 (1-310) x US-09-392-184-31 (1-612)
Qy 127 LeuPheIleTyrglygylProGlnGlyGlnIleGluIleAspArgLysValGluGlyLeu 146
Db 543 GTGCTNATCCGGGNAACN-----NAAAGATTGTGCGATCAGTGGGAGACTC 493
```

```
Qy 147 -GlnTyLeuAlaSerArGTyArSpleuIleAspLeu-AspArgValGlyIleHISGlyT 166
Db 492 CCATATCTAATCTTCTGAAATGATTTTCATTGACNTAAAGATCGTGGCAATCCNCGCT 433
Qy 166 rPserTyrglygylTyTyLeuSerLeuMeAlaLeuMeGlnArGserAspIlePheArGv 186
Db 432 G-TCTATGAGAGATACCTCTCCCTGATGCGCANTAAATTCAGAGCTCAGATATCMTCAAGG 374
Qy 186 AlaIleAlaGlyAlaProValThrLeuPheTyArPThrGlyTyThrGlu 206
Db 378 TTCT-ATTTCGTGGGCCCCAGTCACTGTGATCTTCTAAGAAACAGAAACNCGGAC 315
Qy 206 rGlyrMeGlyYHISProAspGlnAsnGluGlnGlyTyTyTyLeuGlySerValAlaMeG 226
Db 314 GTTAAATGGGTCACTCCNNCCAGAAATGACAGCGCTATTACTTAAGATCTGTGGCAATC 255
Qy 226 lAlaGlyLysPheProSerGluProAsnArGlyLeuLeuHISGlyPheLeuAspG 246
Db 254 CACAGGAAAGTCCCTCTGMAACCAATNNGTTCCTTAAATGATTTCTTGAT 195
Qy 246 lAsnValHISpHeAlaHISThrSerIleLeuLeuSerPheLeuValArGAlaGlyLysP 266
Db 194 AGGATTTCCANTTTTCAACAAACAGNAAATTAATTAAGTTTTHGTAGAGCTGGAAAC 135
Qy 266 rOTyArSpleuGlnIleTyTyProGlnGluArGHisSerIleArGValProGluSerGlyG 286
Db 134 CAATGANTNNCAAGAACCACTCNGAGAAACCAANNAGAGTNTGAAATCGGAG 75
Qy 286 lHisTyTyr 288
Db 74 GACAAATAT 67

RESULT 3
US-10-002-593-5
; Sequence 5, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTII
; FILE REFERENCE: Acty Docet No. 6586198 1242/48/2
; CURRENT APPLICATION NUMBER: US/10/002,593
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/244,524
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3407
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-593-5

Alignment Scores:
Pred. No.: 1.59e-29 Length: 3407
Score: 337.50 Matches: 99
Percent Similarity: 42.94% Conservative: 53
Best Local Similarity: 27.97% Mismatches: 127
Query Match: 20.09% Indels: 75
DB: 4 Gaps: 13

US-10-070-464-3 (1-310) x US-10-002-593-5 (1-3407)
Qy 5 LysAspSerProLeuGlnHISLeuTyValValSerTyValAsnProGlyGlyVal 24
Db 1342 AAAGAAATGCCAGAGAGAGAAATCTTTATTAATCAACTATGATGATCAAAAGTG 1401
Qy 25 ThrArGLeuThrArAspArgGlyTyTySerHISerCysIleSer---GlnHISCyAsp 43
Db 1402 ACATGCTC-----AGTTGTAGCTGAATCCGAAAGGTGTCTAG 1440
```

```

Qy 44 phepelleSerlySerAenglnlyAsnProHis-----Cys----- 57
Db 1441 TACTATTCTGTGTCATCAGTAAGAGGAGAACTATTATCAGCTAGATGTCGGTCC 1500
Qy 58 ---ValSerLeuTyrlYsLeuSerSerProGluAspProThrCyslyThrlyGlu 76
Db 1501 GGCTGGCCCCCTATATCTATACACACCGCTGAATAAAGCGCTGAGCTCGTGA 1560
Qy 77 PheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyThrProGlu--- 95
Db 1561 GAC---AATTCACTTGGATATAAATG-----CTGCAGATGTCAGATGCCCTCAA 1611
Qy 96 -----IlePheSerPheGluSerThrThrGlyPheThrLeuTyrlMetLeuTyrlYs 113
Db 1612 AAACGTGACTCATTTATTTGAATGAACAAAATTT---TGGTATCAGATGATCTTGCT 1668
Qy 114 ProHisAspLeuGlnProGlylySlyTyrlProThValleuPheIleTyrlGlyPro 133
Db 1669 CCTCAT---TTGTATTAATTCAGAAATATCTCTATATGATGTGTAAGAGCCCA 1725
Qy 133 ----- 133
Db 1726 TGTAAGCAAAAAGACACACTGTCTTCAAGTGAAGTGGGCCACTTACCTTGCAAGACA 1785
Qy 133 ----- 133
Db 1786 GAAACATTATAGTACTAGCTTTGATGACAGAGAGTGTACCAAGAGATAGATG 1845
Qy 134 -----GlnGlyGlnIleGlnIleAspAspGlnValGluGlyLeu 146
Db 1846 ATGCATGCATCAACAGAGAACTGGAGACATTTGAAGATGAATGAAGCAGCC 1905
Qy 147 GlnTyrlLeuAlaSerArgTyrlAspPheIleAspLeuAspArgValGlyIleHisGlyTyrl 166
Db 1906 AGACAAATTT---TCAAAAATGGGATTTGGGACAAACAAAGCAATTGCAATTTGGGGCTGG 1962
Qy 167 SerTyrlGlyGlyTyrlLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgVal 186
Db 1963 TCATATGAGAGGTACCTAATGCTGCGGATGCGGAGGAGTGGCTGTTCAAGTGT 2022
Qy 187 AlaIleAlaGlyAlaProValThrLeuThrIlePheTyrlAspThrGlyTyrlThrGluArg 206
Db 2023 GGAATAGCCGTGCGCTGATATCCCGGTGGAGACTATGATGACAGTGAACAGACGT 2082
Qy 207 TyrlMetGly-----HisProAspGlnAenglnGlnGlyTyrlTyrlLeuGlySerValAla 224
Db 2083 TACATGGGTCTCCCAACTCCAGAGCAACCTTGACCATTTACAGAAATTCACACAGTCAG 2142
Qy 225 MetGlnAlaGlyLeuPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 244
Db 2143 AGCAGAGCTGAAATTTT-----AAACAAGTTGAGTACCTCTTATTCATGGAACAGCA 2196
Qy 245 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 264
Db 2197 GATGATTAACGTTCACTTACAGAGCTGACAGATTCCAAAGCCCTGGTGCATGTTGA 2256
Qy 265 LysProTyrlAspLeuGlnIleTyrlProGlnIleArgHisSerIleArgValProGluSer 284
Db 2257 GTGAGTTTCCAGGCAATGTGTAATCTAGTAAGACCAATGGAATGCTAGCAGCAGCA 2316
Qy 285 GlyGluHisTyrlGluLeuHisLeuLeuHisTyrlLeuGlnGlu 298
Db 2317 CACCAACATATATATATACCAATGAGCCACTCATTAACAA 2358

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2924
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US93-07923-1
Alignment Scores:
Pred. No.: 3,72e-29
Score: 333.50
Percent Similarity: 42.18%
Best Local Similarity: 27.65%
Query Match: 19.85%
DB: 5 Gaps: 14
US-10-070-464-3 (1-310) x PCT-US93-07923-1 (1-2924)
Qy 5 LysAspSerProLeuGlnHisIleuTyrlValIleSerTyrlValAsnProGlyGluVal 24
Db 1277 AAAGAAATGCCAGAGAGAAAGAAATCTTATATAAATC----- 1312
Qy 25 ThrArgLeuThrAspArgGlyTyrlSerHisSerCysIleSer----- 39
Db 1313 ---CAACTAGTAC-----TATCAAAAGTACATGCTCAGTGTGAGCTGAATCCG 1363
Qy 40 GlnHisCysAspPheIleSerTyrlSerAenglnlyAsnProHis----- 56
Db 1384 GAAAGGTGCTAGTATATCTGTGCTATCAGTAAGAGCGCAAGTATATATACGTCGACA 1423
Qy 57 Cys-----ValSerLeuTyrlYsLeuSerSerProGluAspAspProThrCys 72
Db 1424 TGTTCCGGTCTGCTGCTGCCCTTATATCTTACACAGAGCGTGAATGAAGGCGCTG 1483
Qy 73 LysThrlyGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyThr 92
Db 1484 AGAGTCTGGAAGAC---AATTCACTTGGATATAAATG-----CTGCAGATGTCGAG 1534
Qy 93 ProProGlu-----IlePheSerPheGluSerThrThrGlyPheThrLeuTyrlGly 109
Db 1535 ATGCCCTCCAAAACCTGCACTTCTATTATTTGAATGAACAAAATTT---TGGTATCAG 1591
Qy 110 MetLeuTyrlYsProHisAspLeuGlnProGlylySlyTyrlProThValleuPheIle 129

```



```

Db      1592 ATGATCTTGCCTCCATCAT---TTTGATTAATCCAGAAATATCTTACTATTAGATG 1648
Qy      130 TYRGLYGLYPR-----
Db      1649 TATGACAGGCCCATGTAGTCAAAAAGACACGTCTTCAAGTGAACCTGGCCACTTAC 1708
Qy      133 -----
Db      1709 CTTCGACAGCAGAAAAACATTATAGTAGCTTGTGATGCGAGAGAGTGTACCAA 1768
Qy      134 -----
Db      1769 GGAATATAGATCATGATGATGCAACAGACAGCTGGAACATTGTAAGTGAATCA 1828
Qy      143 VALGLUGLYLEUANTYRLEUVALSERARGTYRASPHEILEASPLEUSPARVALGLY 162
Db      1829 ATTTGAGCAGCCAGACAAATTT---TCAAAATGGGATTTGTGACAAACAAACAAATGCA 1885
Qy      163 HLEHSGLYTRPSERTYRGLYGLYTYRLEUSERLEUMETALAEUMETGLNARGSERAP 182
Db      1886 ATTTGGGGCTGGTCATATGAGAGGTACGTAAACCTCATATGCTCTGGGATCAGAGAGTGGC 1945
Qy      183 HLEPHARGVALALALALAGLYALAPROVALTHREUTRIPLEPHEITYRASPTRHGLY 202
Db      1946 GTGTTCAAGTGTGGAATAGCCGCGGCTGTATCCCGTGGAGTAGTACTATGACTAGT 2005
Qy      203 TYRTHGLUARGTYRMEGLY-----HISPROASPDNLANSGLINGLYTYRGLY 220
Db      2006 TACACGAAAGCTTACATAGGGCTCCCAACTCCAGAAACAACTTGACCTTACAAAT 2065
Qy      221 GLYSERVALALAMETLINALGLIULYSPHEPROSEGLUPROASNRGLLEULEULEU 240
Db      2066 TCAACAGTCATGACGAGCTGAAAATTTT---AAACAGTTGAGTACCTCTTAT 2119
Qy      241 HISGLYRHELEUSPDLSNVALHISPHLEALHISHTHRSERILEULEUSERPHELU 260
Db      2120 CATGGAACAGCAGATATATACGTTCACTTCCAGACAGTCAAGTCTCCAAAGCCCTG 2179
Qy      261 VALARGALGLYLYBPROTYRASPHEUGLILETYRPROGLINGLYRGLHISERILEARG 280
Db      2180 GTGATGTTGGAGTGTGATTTCCAGGCATGTGATATGATGATGAACCATGGAATAGCT 2239
Qy      281 VALPROGLUSERGLYHISHTYRGLIULYHISLEULEUHIESTYRLEUGLINGLY 298
Db      2240 AGCAGACAGCAGACCAACATATATATACCCACATAGCCACTTCAAAAACAA 2293

RESULT 5
US-09-462-284-1
; Sequence 1, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Affolter, Michael
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5496
; TYPE: DNA
; ORGANISM: Fungus
US-09-462-284-1

Alignment Scores:
Pred. No.: 5,79e-28 Length: 5496
Score: 327.00 Matches: 103

```

```

Percent Similarity: 41.45% Conservative: 40
Best Local Similarity: 29.86% Mismatches: 126
Query Match: 19.46% Indels: 76
DB: 4 Gaps: 11

US-10-070-464-3 (1-310) x US-09-462-284-1 (1-5496)

Qy      4 THRLEASPSERPROLEUGLHISHTYRGLYVALSERTYRVALASNPGLYGLY 23
Db      3146 ACTCAACACCAAGCAGACCGGAGGCTCTCTACCTCGTCTCTAT---TCCAGTTTGGC 3202
Qy      24 VALTHRARGLEUTHASPARGLYTYRSE-----HISERCYSGLILESGINHS 41
Db      3203 GTACCCCGCTCGTGCACACACCGTTCGGGCTACTGCTTCTTCTTCCGGGAGC 3262
Qy      42 CYASRPHRPHLEISERTYRSELYSERLEANGILYASNPRIHISCYVALSERLEUY 61
Db      3263 TCGGGCTACTACATCTCTCATATCGAGGCCCCAGACGTACCTTAC---CAGAACCTTAC 3319
Qy      62 LYSLEUSERPROGLIUSPARSPROTHRCYLYSTHRYGLUPHEITRPLATHRIE 81
Db      3320 ACAGCCACAGT-----ACCAACCACTC---GCACACATC 3352
Qy      82 LEUASPSERLALGLYPROLEU-----PROASPTYRTHPROGLIUPHEISER 98
Db      3353 ACCGACAAAGCCAAAGTACTCGACGAATCAAGAGTATGATTGCCCAACATCACCTAC 3412
Qy      99 PHEGLY-----SETHRTHGLYPHEITRLEUYTYRGLYMETLEUYRILYSPRHSAP 116
Db      3413 TTGAGCTTCCCTCCCTCCCGAGAAACCTCAATGTATGATGACAGGCTTACCCCGGG 3472
Qy      117 LEUGLNPGLIYLYLYSTYRPROTHRYALAEUPHEITRYGLYGLYPROGLINGLY 136
Db      3473 TTCTCCCGATTAAGATGATCCCATCTTTTACCCCATACGCGGCCGAGCCGCA 3532
Qy      136 -----
Db      3533 GAAGTGACCAAGATGCGAAGCCTGAATTCAAGCCTATGTGCTCCGACAGCGAA 3592
Qy      136 -----
Db      3593 CTGAGTACTGTAACCTGACCTGCAACAACCGCGACAGTTTCAAGAGCCCAAGTTC 3652
Qy      137 -----
Db      3653 CCTCGCGCGTACGCGCAACTCGGCTCGCTCGTGAAGCAGAAACCATGCTACGCGCG 3712
Qy      147 GLINTYRLEUALSERARGTYRASPHEILEASPLEUSPARVALGLYHISGLYTR 166
Db      3713 CAACAG--GCGGCCAAACATCCCTGGATGTGACGACACATCGGCATCTGGGGCTG 3769
Qy      167 SETYRGLYGLYTYRLEUSERLEUMETALAEUMETGLNARGSERAPLIEPHEARGVAL 186
Db      3770 AGTTTGGAGCTACTGTGACGAAAGCTCTCGAAGAGACAGCGGCTTTCACATTA 3829
Qy      187 ALALIEALGLYALAPROVALTHREUTRIPLEPHEITYRASPTRHGLYTYRTHGLUARG 206
Db      3830 GGAAGTCATACCGCCCTTTCTGATCGCTTCTTCAAGTCAATGATGACAGGAGCC 3889
Qy      207 TYRMEGLYHISPROASPDLSNGLINGLYTYRGLYGLYSERVALAMETGLIN 226
Db      3890 TACATGAAAGACCTTCCAGCAATAGAGGGCTCGAGACCAAGCCGCTC---CGCAAG 3946
Qy      227 ALAGIULYRPHROSEGLIUPROASNRGLLEULEULEUHIESTYRLEUASPDGIN 246
Db      3947 ACTGACGGGTTCAAGAACGTGAGGGCGGATCTTGATTCACACAGGAGCGGCGAGCAT 4006
Qy      247 AENVALLHSPHEALAHISHTHRSERILEULEUSERPHELEUVALARGALGLYLYBPRO 266
Db      4007 AACGTCATTTCCAGAACTCGGCTCGCTGTGATCTCTGATGCGGCGATGCG----- 4060
Qy      267 TYRASPHEUGLILETYRPROGLINGLYRGLHISERILEARGVALPROGLUSERGLYGLY 286

```

Db 4061 -----GTCCTCCTGAGAACTCCATTGCGAAATGTTCAAGACTGACAC 4108  
Qy 287 HistyrluLeuHis 291  
Db 4109 GGAATCAGCTACCAT 4123

RESULT 6

US-08-230-491A-1  
; Sequence 1, Application US/08230491A  
; Patent No. 5587299  
; GENERAL INFORMATION:  
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;  
; APPLICANT: Gatin-Cheea, Pilar; Old, Lloyd J.  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR  
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: FELFE & LYNCH  
; STREET: 805 THIRD AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERFECT - ASC II  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/230,491A  
; FILING DATE: 20-APRIL-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, NO. 5587299man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 330  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ. ID NO. 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2815 Base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-230-491A-1

Alignment Scores:

Pred. No.: 1,866-25 Length: 2815  
Score: 302.00 Matches: 93  
Percent Similarity: 40.55% Conservative: 55  
Best Local Similarity: 25.48% Mismatches: 139  
Query Match: 17.98% Indels: 78  
DB: 1 Gaps: 11

US-10-070-464-3 (1-310) X US-08-230-491A-1 (1-2815)

Qy 1 PhegluGluYThrIlyAspSerProleuGluHisIleuYrValValSerYrValaAn 20  
Db 1409 TTCACAGTAACACAGATTCAGTTTATCTAGCAATGAATTTGAGAAATACCTGGA 1468  
Qy 21 ProgluGluValThrArgLeuThrAspArgGlyTyrSerHisSer---CysCysIleSer 39  
Db 1469 AGAAGAAACATCTACAGAAATTCATGAGTTCAGAGCTATCTCCAAAGCAAGAGTGTACT 1528  
Qy 40 GluHis-----CysAspPhePheIleSerIlySerAsnGluYsaAn 54  
Db 1529 TCCCATCTTAGAGAAAGAGGCGCAATATTAACAGCAAGTTTACGACATACGCGAAG 1588  
Qy 55 ProHisCysValSerIleuYr-----LysLeuSerSerProGluAspAspPro 70  
Db 1589 TACTATGACCTTGTCTGCTAGCGCCAGGACATCCCATTTCCACCCCTTCATGATGAGACG 1648

Qy 71 ThrCysIleThrIyegluphetrPalatrHisIleAspSerIleaglYProleuProAsp 90  
Db 1649 ACTGATCAAGAA-----ATTAAATCTCTGAAAGAAACAGAAATTTGAAAT 1696  
Qy 91 -----TyrThrProProGluIlePheSerPhegluSerThrIlyPhe 105  
Db 1697 GCTTGAATAATATCCAGCTGCTAAAGAGAAATTAAGAACTTGAAGATGAATAAT 1756  
Qy 106 ThrIleuYrGlymeleuYrIlyProHisIleAspLeuGlnProGlyIlyIlySerProHis 125  
Db 1757 ACTTTATGGTACAGAGTGAATCTCTCTCTCAATTTGACAGATTCAGAAAGATATCCCTTG 1816  
Qy 126 ValIleuPheIleTyrGlyIlyPro----- 133  
Db 1817 CTAATTCAGAGTATAGTGATGCTCCCTGCAATGTAAGTGTATTTCTGTTAAT 1876  
Qy 133 ----- 133  
Db 1877 TGGATATCTTATCTTGAAGTAAGAGGATGTCATTCCTGCTGAGATGTCAGAGA 1936  
Qy 134 -----GlnGlyIleGlu 138  
Db 1937 ACAGCTTTCAGAGTGACAACTCTATGACAGTGTATGAAAGCTGGTGTATTATGA 1996  
Qy 139 IleAspAspGluValIleGluIleuGlnIlyleuAlaSerArgTyrAspPheIleAspLeu 158  
Db 1997 GTTGAAGACCAAGATTCACAGCTGTCAGAAATTCATTA---GAAATGGCTTCATGTATGA 2053  
Qy 159 AspArgValIlyIleHisIleGlyTyrPheSerTyrGlyIlyTyrLeuSerIleuAlaIleuMet 178  
Db 2054 AAAAGAAATGCATATGAGGCGTGTCTTATGAGATATGTTTACATGAGCTTGTGA 2113  
Qy 179 GluArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrIleuTrpIlePhe 198  
Db 2114 TCTGAACTGCTCTTTCAAAGTGTATGAGAGTGGCTCCAGTCTCCAGCTGGGAATAT 2173  
Qy 199 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro-----AspGlnAsnGluGln 216  
Db 2174 TACGGCTCTGTCTACACAGAGATTCATGAGGCTCTCCCAAGAGATATATATCTTGAG 2233  
Qy 217 GlyTyrTyrLeuGlySerValAlaIleGlnAlaGlyIlyPheProSerGluProAsnArg 236  
Db 2234 CACTATTAAGAAATTCACAGTGTATGAGAGAGAGAGAAATTTTCAAGAAATGATAGC----- 2287  
Qy 237 LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleu 256  
Db 2288 TATCTTTCATCCAGGAAACAGCAGATGATATATGTCACCTTCAAAAACACAGCAGATT 2347  
Qy 257 LeuSerPheLeuValArgAlaGlyIlyProTyrAspLeuGlnIleTyrProGlnGluArg 276  
Db 2348 GCTAAAGCTCTGCTTAATGACAGAGTGTATTCAGGCAATGTGTACTGTACAGAAC 2407  
Qy 277 HisSerIleArgValProGluSerGly-----GluHisIyGluLeuHisIleu 293  
Db 2408 CACGGCTTA-----TCCGCGCTGTCCAGAAACCTTATTAACCCACATGACC 2455  
Qy 294 HistyrluGlnGlu 298  
Db 2456 CACTTCTTAAGCAG 2470

RESULT 7

US-08-619-280A-1  
; Sequence 1, Application US/08619280A  
; Patent No. 5767242  
; GENERAL INFORMATION:  
; APPLICANT: Zimmermann, Rainer; Park, John E.;  
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.  
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN  
; TITLE OF INVENTION: ALPHA, AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: FELFE & LYNCH

STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/619,280A  
FILING DATE: 18-MARCH-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/230,491  
FILING DATE: 20-APRIL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5767242man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5330.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2815 Base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-619-280A-1  
Alignment Scores:  
Pred. No.: 1,86e-25 Length: 2815  
Score: 302.00 Matches: 93  
Percent Similarity: 40.55% Conservative: 55  
Best Local Similarity: 25,488 Mismatches: 139  
Query Match: 17,988 Indels: 78  
DB: 1 Gaps: 11  
US-10-070-464-3 (1-310) x US-08-619-280A-1 (1-2815)  
QY 1 PheGluGlyThrIysAspSerProLeuGluHisIleuLeuTyValValSerTyValaAn 20  
DB 1409 TTCAGAGTACACAGAGATTCACGTTTATTTACAGCAATGAATTTGAAGATCCCTGGA 1468  
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTyTserHisSer---CysCysIleSer 39  
DB 1469 AGAAGAAACATCTACAGAAATTTACAGATTCGAAAGCTATCTCCAGCAAGAAAGTGTGTACT 1528  
QY 40 GlnHis-----CysAspRhePheIleSerIysTyTserLengInLysaAn 54  
DB 1529 TGGCACTTACAGAAAGAGTCCCAATATTACACAGCAAGTTTCAGCGCACTACGCCAAG 1588  
QY 55 ProHisCysValSerLeuTy-----LysLeuSerSerProGluAspArgPro 70  
DB 1589 TACTATGACACTTCTGCTACGCGCCAGGATCCCATTTCCACCTTATATATGAGACGC 1648  
QY 71 ThrCysLeuThrIysGluPheThrAlaThrIleLeuAspSerXlaGlyProLeuProAsp 90  
DB 1649 ACTGATCAAGAA-----ATTAAATCTCTGAGAAAGAAACAAAGAAATTTGAAAAAT 1696  
QY 91 -----TyThrProProGluIlePheSerPheGluSerThrArgIlyPhe 105  
DB 1697 GCTTTAAAAATATCCAGCTGCTTAAAGAGAAATTTAAGAACTTAAGATGATAAT 1756  
QY 106 ThrLeuTyGlyMetLeuTyIlysProHisAspLeuGlnProGlyIysIlyTyProThr 125  
DB 1757 ACTTATGACAGCAAGATGATTTCTCTCCATTTGACAGATCAAGAAAGATATCCCTTG 1816  
QY 126 ValLeuPheIleTyArgIlyPro----- 133  
DB 1817 CTAAATCAAGTATATGTGTGTCTCTGCAAGTCAAGTGAAGTGTATTTGCTGTAAAT 1876

QY 133 ----- 133  
DB 1877 TGGATATCTTATCTTCAAGTAAGAGGATGTCATTCCTGTGGATGTCGAGCA 1936  
QY 134 -----GlnGlyGlnIleGlu 138  
DB 1937 ACAGCTTTCCAAAGTCAAACTCTCTATGACGTATGCAAGCTGGGTGTTATGAA 1996  
QY 139 IleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspRheIleAspLeu 158  
DB 1997 GTTGAAGACCAAGTTTACAGTGTACAGAAATTCATA---GAATGGGTTTCATTTGAA 2053  
QY 159 AspArgValGlyIleHisGlyTyrPserTyArgIlyTyTleuSerLeuMetAlaLeuMet 178  
DB 2054 AAAAGATAGCCATATGAGGAGTGTCTCTATGAGAGATAGCTTTCATCAGCTGGCCTTGA 2113  
QY 179 GlnArgSerAspRheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePhe 198  
DB 2114 TCTGAACTGTCTTTTCAATGTGTATAGCAAGTGTCCAGCTTCCAGCTGGAAATAT 2173  
QY 199 TyAspThrGlyTyThrGluArgTyMetGlyHisPro-----AspGlnAsnGluGln 216  
DB 2176 TACGCTGTCTGTACACAGAGATTCATGCGTCTCCAAAGATGATATCTTGAG 2233  
QY 217 GlyTyTyrLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArg 236  
DB 2234 CACTATAGAAATTCAACTGTGATGCGCAAGAGCAAGATATTTCAGAAATGTAGAC----- 2287  
QY 237 LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 256  
DB 2288 TATCTTCTATCCACGAGAACAGCAGATGATTAATGTACCTTTCAAAGCTCAGCAGAT 2347  
QY 257 LeuSerPheLeuValArgAlaGlyLysProTyAspLeuGlnIleTyTyrProGlnGluArg 276  
DB 2348 GCTAAAGCTGTGTATATGACAAAGATGATTTCCAGCAATGCGTATCTGACCAAGAC 2407  
QY 277 HisSerIleArgValProGluSerGly-----GlnHisTyGluLeuHisLeuLeu 293  
DB 2408 CACGGCTTA-----TCCGCGCTGTCCACGCAACCACTTATATACCCACATGACC 2455  
QY 294 HisTyLeuGlnGlu 298  
DB 2456 CACTTCTTAAAGCAG 2470  
RESULT 8  
US-08-940-391-1  
Sequence 1, Application US/08940391  
Patent No. 5965373  
GENERAL INFORMATION:  
APPLICANT: Zimmermann, Rainer; Park, John E.;  
TITLE OF INVENTION: Ret(19, Wolfgang; Old, Lloyd J.  
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/940,391  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/619,280

```

/ FILING DATE: 18-MARCH-1996
/ APPLICATION NUMBER: 08/230,491
/ FILING DATE: 20-APRIL-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hanson, No. 5965373man D.
/ REGISTRATION NUMBER: 30,946
/ REFERENCE/DOCKET NUMBER: LUD 5330.1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 688-9200
/ TELEFAX: (212) 638-3884
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2815 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/
US-08-940-391-1

Alignment Scores:
Pred. No.: 1,866-25 Length: 2815
Score: 302.00 Matches: 93
Percent Similarity: 40.55% Conservative: 55
Best Local Similarity: 25.48% Mismatches: 139
Query Match: 17.98% Indels: 78
DB: 2 Gaps: 11

US-10-070-464-3 (1-310) x US-08-940-391-1 (1-2815)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuValValSerTyrValaen 20
DB 1409 TTCAGAGTACACAGAGTCACTGTTTATTCTAGCAATGAATTTGAGAAATACCTGGA 1468
QY 21 ProGluValThrArgLeuThrAspArgGlyTyrSerHisSer---CysCysHisSer 39
DB 1469 AGAAGAACATCTACAGATTCAGATTGAGAGTATCCCAAGAGAGAGTGTACT 1528
QY 40 GluHis-----CysAspPhePheHisSerLysTyrSerAsnGlnLysAsn 54
DB 1529 TGCCCATCTAAGAGAAAGAGGTCGCAATATTACACAGCAAGTTTCAGCACTACGCAAG 1588
QY 55 ProHisCysValSerLeuTyr-----LysLeuSerSerProGluAspAspPro 70
DB 1589 TACTATGCACTGTCTGCTACGAGCCAGGCAATCCCATTTCCACCTTCATATGAGACGC 1648
QY 71 ThrCysLysThrLysGluPheThrPalaThrLysLeuAspSerAlaGlyProLeuProAsp 90
DB 1649 ACTGATCAAGAA-----ATTAAATCCCTGGAAGAAACAAGAAATGGAAT 1696
QY 91 -----TyrThrProProGluLysPheSerPheGluSerThrGlyPhe 105
DB 1697 GCTTGAATAATATCCAGCTGCTAAGAGAAATTAAGAACTTGAAATGATGAAT 1756
QY 106 ThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThr 125
DB 1757 ACTTATAGGTACAAAGATGATTTCTTCTCTCAATTGACAGATCAAGAAAGATCCCTTG 1816
QY 126 ValLeuPheLeuTyrGlyPro----- 133
DB 1817 CTAATTCAAGGTATAGTGTCTCCGACATGACAGATGAAGTCTGATTTGCTGTAAT 1876
QY 133 ----- 133
DB 1877 TGGATATCTTATCTTGAAGTAAGAGGATGTCATTCCTGTGAGATGTCAGAGA 1936
QY 134 -----GluGlyGlnLysGlu 138
DB 1937 ACAAGTTTCAAGGTGACAAACTCTCTATGACAGTGTATCGAAAGCTGGGTTTATAGA 1996
QY 139 IleAspArgGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeu 158
DB 1997 GTTGAAGACCAAGATTAAGAGTGTCAAGAAATTCATA---GAATGGGTTTCAATTGATGA 2053
QY 159 AspArgValGlyIleHisGlyTyrPheSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMet 178

```

```

DB 2054 AAAAGATAGCCATATGGGGCTGCTCTATGAGAGATACGTTTCACTAGCGCCCTTGA 2113
QY 179 GlnArgSerAspIlePheArgValAlaIleAlaProValThrLeuThrIlePhe 198
DB 2114 TCTGAAGCTGCTTTTCAAAAGTGTATAGAGTGGCTCCAGTCTCCAGCTGGAAAT 2173
QY 199 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro-----AspGlnAsnGluGln 216
DB 2174 TACGGCTCTGTACACAGAGATTCATGAGTCTCCCAAGAAATATATATCTTGA 2233
QY 217 GlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArg 236
DB 2234 CACTATTAAGAAATTAACCTGATGCGACAGACAGAAATATTTCAGAAATATAGAC----- 2287
QY 237 LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 256
DB 2288 TATCTTCATCCAGGAACAGACAGATGATATATGTCACCTTCAAAATCAGACAGAT 2347
QY 257 LeuSerPheLeuValAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArg 276
DB 2348 GCTAAAGCTCTGTTAATGCAACAAGTATTCACAGCAATGTGTACTTACCAAGAAC 2407
QY 277 HisSerIleArgValProGluSerGly-----GluHisTyrGluLeuHisLeu 293
DB 2408 CACGCTTA-----TCCGAGCTGTCCAGAACCACTTATACACCCACATGACC 2455
QY 294 HisTyrLeuGlnGlu 298
DB 2456 CACTTCTTAAGCAG 2470

RESULT 9
US-08-699-103B-1/c
/ Sequence 1, Application US/08699103B
/ Patent No. 6107462
/ GENERAL INFORMATION:
/ APPLICANT: Hampe, Jasper D.
/ TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
/ TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Fish & Richardson P.C.
/ STREET: 2200 Sand Hill Road, Suite 100
/ CITY: Menlo Park
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94025
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/699,103B
/ FILING DATE: 16-AUG-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/002,581
/ FILING DATE: 17-AUG-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Green, Grant D.
/ REGISTRATION NUMBER: 31,259
/ REFERENCE/DOCKET NUMBER: 09272/005001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650/322-5070
/ TELEFAX: 650/854-0875
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4982 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA

```

US-08-699-103B-1

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
6.19e-21	267.00	42.65%	31.75%	3	4982	67	23	61	60	6

US-10-070-464-3 (1-310) x US-08-699-103B-1 (1-4982)

```

QY 123 TyrProthValLeuPheIleTyrglyProgin-----134
DB 4941 TATCCTGATTTTCTTTCATATGCGGAGCCGAATTCACAAAGTTGTCAAAACGTTT 4882
QY 134 -----134
DB 4881 TCCGTAGATTATATGAAGTGTAGCTTCACAATTAAAGCAATTGAGTTGTTGAC 4822
QY 135 -----gly 135
DB 4821 GGTGCTGACTGCTTCAAAAGTCAAGACTTAGATCCCTTGTTCGATAGCTCGGT 4762
QY 136 GlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspPhe 155
DB 4761 GATTACGAGCGCCGCGACCAAAATATCTCGCGCTTCTTATATGTTCT--TTAAGTTT 4705
QY 156 IleAspLeuAspArgValGlyIleHisGlyTyrSerTyrglyTyLeuSerLeuMet 175
DB 4704 GTTGATCCGCAAAAGTTTCTTATTTGTTGTCATACGGGGGTACCTGACACTAAAA 4645
QY 176 AlaLeuMetGlnArgSerAsp--IlePheArgValAlaIleAlaGlyAlaProValThr 194
DB 4644 ACTTTGAGAAAGATGCGGAGAACATTCCTCAATACGGATGTCAGTTCCGCGAGTAACC 4585
QY 195 LeuTyrIlePheTyAspThrGlyTyrThrGluArgTyMeGlyHisProAspGlnAsn 214
DB 4584 GACTGAGACTTTTACATTCCTTTTACTGAGAGGATACATACCTCTCAAGAAAA 4525
QY 215 GluGlnGlyTyrTyLeuGlySerVal-----AlaMetGlnAlaGluAspPhe 230
DB 4524 TTTGATGATACGTATGATCAACGCTTCAATATGTCAGCTTTG-----4480
QY 231 ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPhe 250
DB 4479 ---GCAACAAGCAAAATGATTTTGTGATGACACGGAACAGAGATGATTAACGTTCACTT 4423
QY 251 AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly---LysProTyAspLeu 269
DB 4422 CAATATCCCTAAAGTTTCTGACCTTTTGATCTAAATGATGTGGAAAAATATATGACGTC 4363
QY 270 GlnIleTyProGlnGluArgHisSerIleArg 280
DB 4362 CACGCTTCTCTGACTCAGATCATATATTAAGA 4330

```

RESULT 10

US-09-229-059-1/c

Sequence 1, Application US/09229059

Patent No. 6333172

GENERAL INFORMATION:

APPLICANT: Rine, Jasper D.

APPLICANT: Hampton, Randolph

TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING

NUMBER OF SEQUENCES: 25

CHOLESTEROL SYNTHESIS

ADDRESS: Fish & Richardson P.C.

STREET: 2200 Sand Hill Road, Suite 100

CITY: Menlo Park

STATE: CA

COUNTRY: USA

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/229,059

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/599,103

FILING DATE: 16-AUG-1996

APPLICATION NUMBER: 60/002,581

FILING DATE: 17-AUG-1995

ATTORNEY/AGENT INFORMATION:

NAME: Green, Grant D.

REGISTRATION NUMBER: 31,259

REFERENCE/DOCKET NUMBER: 09272/005001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/322-5070

TELEFAX: 650/854-0875

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4982 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-09-229-059-1

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
6.19e-21	267.00	42.65%	31.75%	4	4982	67	23	61	60	6

US-10-070-464-3 (1-310) x US-09-229-059-1 (1-4982)

```

QY 123 TyrProthValLeuPheIleTyrglyProgin-----134
DB 4941 TATCCTGATTTTCTTTCATATGCGGAGCCGAATTCACAAAGTTGTCAAAACGTTT 4882
QY 134 -----134
DB 4881 TCCGTAGATTATATGAAGTGTAGCTTCACAATTAAAGCAATTGAGTTGTTGAC 4822
QY 135 -----gly 135
DB 4821 GGTGCTGACTGCTTCAAAAGTCAAGACTTAGATCCCTTGTTCGATAGCTCGGT 4762
QY 136 GlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspPhe 155
DB 4761 GATTACGAGCGCCGCGACCAAAATATCTCGCGCTTCTTATATGTTCT--TTAAGTTT 4705
QY 156 IleAspLeuAspArgValGlyIleHisGlyTyrSerTyrglyTyLeuSerLeuMet 175
DB 4704 GTTGATCCGCAAAAGTTTCTTATTTGTTGTCATACGGGGGTACCTGACACTAAAA 4645
QY 176 AlaLeuMetGlnArgSerAsp--IlePheArgValAlaIleAlaGlyAlaProValThr 194
DB 4644 ACTTTGAGAAAGATGCGGAGAACATTCCTCAATACGGATGTCAGTTCCGCGAGTAACC 4585
QY 195 LeuTyrIlePheTyAspThrGlyTyrThrGluArgTyMeGlyHisProAspGlnAsn 214
DB 4584 GACTGAGACTTTTACATTCCTTTTACTGAGAGGATACATGATCAATCCTCAAGAAAA 4525
QY 215 GluGlnGlyTyrTyLeuGlySerVal-----AlaMetGlnAlaGluAspPhe 230
DB 4524 TTTGATGATACGTATGATCAACGCTTCAATATGTCAGCTTTG-----4480
QY 231 ProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPhe 250
DB -----4480

```

Db 4479 ---GCACAAAGCAATAGATTGTTGGATGACGAGAAAGAGATGATAACGTTCACTTT 4423

Qy 251 AlAHistSerileuSerPheLeuValArgAlaGly---LysProTyrAaPleu 269

Db 4422 CAAATTCCTTAAGATTCTTGACCTTTTGATCTAAATGCTGAGAAAATATGACGTC 4363

Qy 270 GlnIleTyrProGlnGluArgHisSerileArg 280

Db 4362 CACGCTCTTCTGACTCAGATCATAGATATAGA 4330

RESULT 11

US-09-628-133-1/c

Sequence 1, Application US/09628133

Patent No. 6531292

GENERAL INFORMATION:

APPLICANT: Rine, Jasper D.

TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 2200 Sand Hill Road, Suite 100

CITY: Menlo Park

STATE: CA

COUNTRY: USA

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/628,133

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/699,103

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Green, Grant D.

REGISTRATION NUMBER: 31,259

REFERENCE/DOCKET NUMBER: 09272/005001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/854-0875

TELEFAX: 650/854-0875

INFORMATION FOR SEQ. ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4982 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-09-628-133-1

Alignment Scores:

Pred. No.: 6.19e-21 Length: 4982

Score: 267.00 Matches: 67

Percent Similarity: 42.65% Conservative: 23

Best Local Similarity: 31.75% Mismatches: 61

Query Match: 15.89% Indels: 60

DB: 4 Gaps: 6

US-10-070-464-3 (1-310) x US-09-628-133-1 (1-4982)

Qy 123 TyrProThValLeuPheIleTyrGlyGlyProGln----- 134

Db 4941 TATCCGTGATTTTCTTGCAATATGGGAGCCGAATTCACAAAGTTGCAAAAGCTTT 4882

Qy 134 ----- 134

Db 4881 TCCGTAGATTAATGAGGTGAGCTTCACATTAAGCAATTTAGTTGTTGAC 4822

Qy 135 -----Gly 135

Db 4821 GGTGCTGTAAGTCTGCTTCAAGGTCGAAGACTTATAGATCCCTTGTCCGATAGGCTCGGT 4762

Qy 136 GlnIleGluIleAspArgIleValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPhe 155

Db 4761 GATTACGAGGCCCGCAGCAAAATATATCTCGGCTTCTTATATAGCTTCT---TTAACTTTT 4705

Qy 156 IleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyrLeuSerLeuMet 175

Db 4704 GTTGATCCCGCAAAAGATTTCTTATTTGGTGTCTATACGGGGGTACTCGACACTATAA 4645

Qy 176 AlaLeuMetGlnArgSerAsp---IlePheArgValAlaIleAlaGlyAlaProValThr 194

Db 4644 ACTTGAGAAAGATGCGGAGACATTTCAATATACGGATGACAGTTGCGCGCAGTAAC 4585

Qy 195 LeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsn 214

Db 4584 GACTGAGATTTTATCATTTCTGTTTATCTGAGAGGTATCATCTCTCAAGAAAC 4525

Qy 215 GluGlnGlyTyrTyrLeuGlySerVal-----AlaMetGlnAlaGlyLysPhe 230

Db 4524 TTGTGATGATACCTAGATCAAGCGTTCTATATGTCAGTCTTGTG----- 4480

Qy 231 ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 250

Db 4479 ---GCACAAAGCAATAGATTGTTGGATGACGAGAAAGAGATGATAACGTTCACTTT 4423

Qy 251 AlAHistSerileuSerPheLeuValArgAlaGly---LysProTyrAaPleu 269

Db 4422 CAAATTCCTTAAGATTCTTGACCTTTTGATCTAAATGCTGAGAAAATATGACGTC 4363

Qy 270 GlnIleTyrProGlnGluArgHisSerileArg 280

Db 4362 CACGCTCTTCTGACTCAGATCATAGATATAGA 4330

RESULT 12

US-09-221-017B-253

Sequence 253, Application US/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FORSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H

```
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 253:
SEQUENCE CHARACTERISTICS:
LENGTH: 543 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...543
US-09-221-017B-253

Alignment Scores:
Pred. No.: 6.01e-22 Length: 543
Score: 263.00 Matches: 51
Percent Similarity: 58.99% Conservative: 31
Best Local Similarity: 36.69% Mismatches: 53
Query Match: 15.65% Indels: 4
DB: Gaps: 2

US-10-070-464-3 (1-310) x US-09-221-017B-253 (1-543)
QY 135 GYGLHLLGLLLEGLLEASPAGLNLGLGLYLEGLNLYRLEUMLASERARGTYRASP 154
    |||||
DB 136 GGGCAGACCGAAGTAGGCCGATGATGTGGGTGATTTCTCAAGACCA--TCA 192
QY 155 PHEILASPLEUASPARGLVGLYLIHISGLYTRPSETYRGLYGLYTYRLEUSETLEU 174
    |||||
DB 193 TGGTGATGTCCTGATGAATAGAGTAGATGCTGGAGCTATGCTTATGACTACG 252
QY 175 METALAEUWETGLNAGSERASPILEPHEARGVALALAEALAGLALAPROVALTHR 194
    |||||
DB 253 AATCTGATGCTTACGCGCGCGATGCTTCAAACTCGAGTAGCGCGCGCTGTCTATA 312
QY 195 LEUTRPLEPHEYRSPRTHGLYTYRTHGLUARGTYRMEGLYHISPROASPLASN 214
    |||||
DB 313 GACTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
QY 215 GLUINGLYTYRTRYLEUGLYSERVALALAMETGLNLAGLUYSRHEPROSERGIUPRO 234
    |||||
DB 373 CCCGAAGATACGATGCTGCGCACCTGCTCAAAAGCGCGGTATCTG-----AAA 423
QY 235 ASNARGLEULEUENLEUHSGLYPHELEUASPLUASNVALHISPHLEALHISTHRSE 254
    |||||
DB 424 GGAAGACTTATGCTGATTCATGAGCAGCATGATCGGCTGATGATGAGCAGCATTC 483
QY 255 ILLEULEUSERPHELEUVALARGALAGLYSEPROTYRSPLEUGLNIETYPRO 273
    |||||
DB 484 CTTTCTTGATGCTGCTGGAAGCAGCAGCACTATCTGACTCTTACGCTATCCG 540

RESULT 13
US-09-221-017B-646/C
Sequence 646, Application US/09221017B
GENERAL INFORMATION:
APPLICANT: ROSS, BRUCE C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CA
```

```
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 646:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...657
US-09-221-017B-646

Alignment Scores:
Pred. No.: 1.61e-20 Length: 657
Score: 252.00 Matches: 56
Percent Similarity: 42.78% Conservative: 24
Best Local Similarity: 29.95% Mismatches: 57
Query Match: 15.00% Indels: 50
DB: Gaps: 3

US-10-070-464-3 (1-310) x US-09-221-017B-646 (1-657)
QY 91 TYRTHPROGSLIPLPHESERPHEGLUSERTHRTHGLYRPHETHRLEUYRGLYMET 110
    |||||
DB 566 TACAACCCGAAGAGATTACCATATCAAACTCAATCGGCTTTGAACATGATGCTGG 507
QY 111 LEUTRYLSEPROHISAPLEUGLNPLOGLYLVSELYRTPROTHRVALLLEUPHELE 130
    |||||
DB 506 ATCGTAAGCTTATTTATTTGATCCCTCGCCACTATCTGCTGATGATGATGATGAT 447
QY 131 GLYGLYPROGLNGLYGLN----- 136
    |||||
DB 446 AGCGGTCCCACTCCAGCAGGATTTGATGCTGCTATTCATTCGATTGGGAACACTACT 387
QY 136 ----- 136
DB 386 GCATGAAGATTACGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 327
```

Qy 137 -----11egluileaspaaplnvalglu 144  
Db 326 GAATGCCGCAAGTACCTACATGCAACTCGGTGATTGCAAAAGGATGATGACATACCA 267  
Qy 145 G1yleuG1nTyreua1SerarG1yRasPhe1leasPleaasParg1y1lehis 164  
Db 266 GCGGCGACTGCTGATTA---GACACGCTGCCCTATGATGATGACAGCTCGATCGGATATG 210  
Qy 165 G1yTPSer1yRg1yG1yTyreuaSer1eumetalaleumetGlna1Serasp1lephe 184  
Db 209 GGGTGAGCTATGCGCGCTATACCACTAATGATGTGTGCGGGAATGATACATTC 150  
Qy 185 ArgVala1a1ealag1yAlaprovalThreutrp1lephetyRasPthrg1yTyThr 204  
Db 149 AAAGCGGGGATAGCGGTGCTCTGTCGACACTGCGCTTTTACATTCGATTTACACC 90  
Qy 205 G1uATyTyMetG1yH1asProasp1nabng1ug1yTyTyreug1ySer----- 222  
Db 89 GAACCTTCATGCTGATACACCCMAAGAGAAATGCTTCGATACAAAGATGCTTCCTCTT 30  
Qy 223 ---Vala1ameGlnalaglu 228  
Db 29 GATGTGCAAGCAATTACAA 9  
RESULT 14  
US-09-389-681-428  
; Sequence 428, Application US/09389681A  
; Patent No. 6518237  
; GENERAL INFORMATION:  
; APPLICANT: Yuguil, Jiang  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.470C3  
; CURRENT FILING DATE: 1999-09-02  
; CURRENT APPLICATION NUMBER: US/09/389,681A  
; NUMBER OF SEQ ID NOS: 463  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 428  
; LENGTH: 535  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-389-681-428  
Alignment Scores:  
Pred. No.: 1.09e-15 Length: 535  
Score: 210.00 Matches: 50  
Percent Similarity: 42.77% Conservative: 21  
Best Local Similarity: 30.12% Mismatches: 45  
Query Match: 12.50% Indels: 50  
DB: 4 Gaps: 3  
US-10-070-464-3 (1-310) x US-09-389-681-428 (1-535)  
Qy 114 ProH1asPleaueGlnProG1yLyelystyRProThrValleuPhe1leTyG1yPro 133  
Db 15 CCTCCTCAATTGACAGATCAAGAAAGATATCCCTTGCTAATTCAAGTATATGATGCTCC 74  
Qy 133 ----- 133  
Db 75 TGCAGTCAGAGTGAAGTCTGTATTTCTGTAAATTGATATCTTATCTTCAAGTAAG 134  
Qy 133 ----- 133  
Db 135 GAAGGATGTCATTGCTCTGTCGATGTCGAGAACAGCTTTCCAAAGTACAAATC 194  
Qy 134 -----GlnG1yGln1leG1u1leasParg1nvalgluG1yLeu 146  
Db 195 CTCATGACAGTATCGAAAGCTGGTGTATGAAAGTGAAGACCAAGATTAACAGCTGC 254  
Qy 147 GlnTyreua1aseraG1yRasPhe1leasPleaasParg1y1lehisG1yTrp 166

Db 255 AGAAATTCATA---GAATGGGTTTCATGATGAAAAAAGAAATGCCATATGGGGCTGG 311  
Qy 167 SerTyRg1yG1yTyreuaSer1eumetalaleumetGlna1Serasp1lephearg1y 186  
Db 312 TCTATGAGAGATAGCTTTCATCACTGCGCTTCATCTGATCGAAGCTGCTTTTCAAAATG 371  
Qy 187 Ala1ealag1yAlaprovalThreutrp1lephetyRasPthrg1yTyThrg1y 206  
Db 382 GGTATGACAGTGGCTTCAGTCTCCAGCTGGAATATTCGGCTGCTGTACACAGAGA 431  
Qy 207 TyMetG1yH1asPro-----Asp1nabng1ug1yTyTyreug1ySerVala1a 224  
Db 432 TTCATGGCTTCACCAACAAAGATGATATCTTGACACTATTAAGATTAACATGATG 491  
Qy 225 MetGlnalagluYsphe 230  
Db 492 GCAAGACGCAATATTTCT 509  
RESULT 15  
US-09-620-405B-428  
; Sequence 428, Application US/09620405B  
; Patent No. 6528054  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Yuguil  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Hepler, William T.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER  
; FILE REFERENCE: 210121.470C8  
; CURRENT APPLICATION NUMBER: US/09/620,405B  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 495  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 428  
; LENGTH: 535  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-620-405B-428  
Alignment Scores:  
Pred. No.: 1.09e-15 Length: 535  
Score: 210.00 Matches: 50  
Percent Similarity: 42.77% Conservative: 21  
Best Local Similarity: 30.12% Mismatches: 45  
Query Match: 12.50% Indels: 50  
DB: 4 Gaps: 3  
US-10-070-464-3 (1-310) x US-09-620-405B-428 (1-535)  
Qy 114 ProH1asPleaueGlnProG1yLyelystyRProThrValleuPhe1leTyG1yPro 133  
Db 15 CCTCCTCAATTGACAGATCAAGAAAGATATCCCTTGCTAATTCAAGTATATGATGCTCC 74  
Qy 133 ----- 133  
Db 75 TGCAGTCAGAGTGAAGTCTGTATTTCTGTAAATTGATATCTTATCTTCAAGTAAG 134  
Qy 133 ----- 133  
Db 135 GAAGGATGTCATTGCTCTGTCGATGTCGAGAACAGCTTTCCAAAGTACAAATC 194  
Qy 134 -----GlnG1yGln1leG1u1leasParg1nvalgluG1yLeu 146  
Db 195 CTCATGACAGTATCGAAAGCTGGTGTATGAAAGTGAAGACCAAGATTAACAGCTGC 254  
Qy 147 GlnTyreua1aseraG1yRasPhe1leasPleaasParg1y1lehisG1yTrp 166  
Db 255 AGAAATTCATA---GAATGGGTTTCATGATGAAAAAAGAAATGCCATATGGGGCTGG 311



```
QY 167 SerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgVal 186
Db 312 TCCTATGAGAGATACGTTTCATCATCTGGCCCTTGCACTGGAACCTGCTTTCAAAATGT 371
QY 187 AlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArg 206
Db 372 GGTATATGACGTGGCTCCAGCTCTCCAGCTGGAAATATTACGCGTCTGTACACAGAGAGA 431
QY 207 TyrMetGlyHisPro-----AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 224
Db 432 TTCATGGGCTCTCCCAACAAGGATGATTAATCTTGAGCACTATAGAAATTCAACTGTGATG 491
QY 225 MetGlnAlaGluLysPhe 230
Db 492 GCAAGAGCAGAAATATTTC 509
```

Search completed: October 16, 2003, 03:30:32  
Job time : 70.5672 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - nucleic search, using frame\_p1us\_p2n model

Run on: October 16, 2003, 00:05:32 (Search time 231.77 Seconds  
(without alignments)  
3512.533 Million cell updates/sec

Title: US-10-070-464-3  
Perfect score: 1680

Sequence: 1 FEGTQDPLEHLLVYVSYVN.....HLHLYOENLGSRIALAKVI 310

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10070464/runat\_15102003\_113555\_24924/app.query.fasta\_1.2652  
-DB=Published\_Applications\_NA -QMT=fastcap -SUFFIX=rmpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pcct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10070464 @CGN 1.1 845 @runat\_15102003\_113555\_24924  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:\*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1645.5	97.9	2649	12	US-10-054-776-1 Sequence 1, Appli

2	1645.5	97.9	2649	12	US-10-170-789-39	Sequence 39, Appli
3	1645.5	97.9	2671	10	US-09-976-674-2	Sequence 2, Appli
4	1645.5	97.9	3143	12	US-10-170-789-17	Sequence 37, Appli
5	1644.5	97.9	4576	10	US-09-976-674-20	Sequence 20, Appli
6	1625.5	96.8	4309	10	US-09-976-674-14	Sequence 14, Appli
7	1625.5	96.8	4829	10	US-09-976-674-12	Sequence 12, Appli
8	1358.5	80.9	4685	10	US-09-976-674-22	Sequence 22, Appli
9	1352	80.5	4523	10	US-09-976-674-8	Sequence 8, Appli
10	1093.5	65.1	2617	10	US-09-976-674-4	Sequence 4, Appli
11	1093.5	65.1	4219	10	US-09-976-674-28	Sequence 28, Appli
12	1093.5	65.1	4302	10	US-09-976-674-24	Sequence 24, Appli
13	1086.5	64.7	4180	10	US-09-976-674-36	Sequence 36, Appli
14	1086.5	64.7	4263	10	US-09-976-674-34	Sequence 34, Appli
15	1032.5	61.5	2801	13	US-10-098-881-100	Sequence 100, App
16	1032.5	61.5	3262	13	US-10-098-881-83	Sequence 83, Appli
17	872.5	51.9	4076	10	US-09-976-674-32	Sequence 32, Appli
18	872.5	51.9	4159	10	US-09-976-674-30	Sequence 30, Appli
19	865.5	51.5	4037	10	US-09-976-674-40	Sequence 40, Appli
20	865.5	51.5	4120	10	US-09-976-674-38	Sequence 38, Appli
21	662.5	39.4	561	11	US-09-764-891-877	Sequence 877, App
22	545	32.4	502	11	US-09-918-985-19585	Sequence 19585, A
23	344.5	20.5	4835	10	US-09-917-800A-1570	Sequence 1570, Ap
24	344.5	20.5	4835	14	US-10-165-603-5	Sequence 5, Appli
25	337.5	20.1	3407	12	US-10-423-714-5	Sequence 5, Appli
26	337.5	20.1	3407	13	US-10-002-593-5	Sequence 5, Appli
27	337.5	20.1	3407	14	US-10-165-603-6	Sequence 6, Appli
28	310	18.5	2130	14	US-10-156-761-3131	Sequence 3131, Ap
29	310	18.5	9025608	14	US-10-156-761-1	Sequence 1, Appli
30	302	18.0	2366	12	US-10-101-510-683	Sequence 683, Appli
31	302	18.0	2788	12	US-10-269-909-22	Sequence 22, Appli
32	302	18.0	2814	10	US-09-962-832-108	Sequence 108, App
33	302	18.0	2814	10	US-09-954-456-1148	Sequence 1148, Ap
34	302	18.0	2814	10	US-09-954-531-367	Sequence 367, App
35	302	18.0	2814	12	US-10-101-510-30	Sequence 30, Appli
36	302	18.0	2814	12	US-10-301-822-34	Sequence 34, Appli
37	302	18.0	2814	12	US-09-873-367C-499	Sequence 499, App
38	302	18.0	2814	14	US-10-177-293-135	Sequence 135, App
39	302	18.0	2815	9	US-09-265-606-1	Sequence 1, Appli
40	302	18.0	3138	14	US-10-198-866-13171	Sequence 13171, A
41	302	18.0	3224	12	US-10-240-965-117	Sequence 117, App
42	289	17.2	2812	14	US-10-084-817-24	Sequence 24, Appli
43	289	17.2	4797	11	US-09-764-891-7074	Sequence 7074, Ap
44	277	16.5	2388	12	US-09-870-133-3	Sequence 3, Appli
45	277	16.5	2388	14	US-10-160-501-6	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-10-054-776-1  
Sequence 1, Application US/10054776  
Publication No. US20030165818A1  
GENERAL INFORMATION:  
APPLICANT: Mark Robert Edbrooke  
APPLICANT: Alan Peter Lewis  
TITLE OF INVENTION: NOVEL PROTEIN  
FILE REFERENCE: OG1042US  
CURRENT APPLICATION NUMBER: US/10/054,776  
CURRENT FILING DATE: 2002-01-23  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 2649  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1) .. (2649)  
US-10-054-776-1  
Alignment Scores: 1.47e-209 Length: 2649  
Pred. No.: 1645.50 Matches: 310

Percent Similarity: 86.35% Conservative: 0  
Best Local Similarity: 86.35% Mismatches: 0  
Query Match: 97.95% Indels: 49  
DB: 12 Gaps: 1

US-10-070-464-3 (1-310) x US-10-054-776-1 (1-2649)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLysLeuValValSerValAlaen 20  
DB 1570 TTTGAAGGACCAAGACTCCCTTTAGAGCATCACTCTAGTACGTACGTAACGTAAT 1629  
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40  
DB 1630 CCGAGAGGAGTACCAAGGCTGACGCTGCGCTACTCACTTCTGCGCATCACTCG 1689  
QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60  
DB 1690 CACTGAGACTTTTATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1749  
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalatnr 80  
DB 1750 TACAAGCTATCAAGTCTCAAGATGACCACTTGCAAAACAAGAAATTTGGGCGCAC 1809  
QY 81 IleuAspSerAlaGlyProLeuProAspTyrThrProGluLysPheSerPheGlu 100  
DB 1810 ATTTGGATTACAGAGGCTCTCTCTGACTATACCTCTCCAGAAATTTCTCTTTTGA 1869  
QY 101 SerThrThrGlyPheThrLeuTyrGlyMeLeuTyrLysProHisAspLeuGlnProGly 120  
DB 1870 AGTACTAGCTGATTTATCATTTGATGAGGATGCTCTCAACGCTCATGATCTACAGCTGGA 1929  
QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134  
DB 1930 AAGAAATATCCACTGCTGCTGCTCATATATGGTGGTCTCCAGGTGCAATTGATATAT 1989  
QY 134 ----- 134  
DB 1990 CCGTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTACCTCTCTAGTTATGTGTT 2049  
QY 134 ----- 134  
DB 2050 GTAAGTATGACACAGGGATCCTGTACCGAGGGCTTAATTTGAAGCGCCTTTANA 2109  
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151  
DB 2110 TATAAATGGTCAATAATGAAATTCGATCGAGTGGAGGACTCCAAATATCTAGCTTCT 2169  
QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 171  
DB 2170 CGATATGATTTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2229  
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
DB 2230 CTCTCCCTGATGAGCTTAATGACAGAGTCAATATCTTCAAGGCTGCTTATGAGGATTC 2289  
QY 132 ProValIleLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
DB 2290 CCAAGTCACTCTGAGATCTTCTATGATGATGATGATGATGATGATGATGATGATGATG 2349  
QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231  
DB 2350 GACCGAATGAAACAGGGCTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 2409  
QY 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
DB 2410 TCTGAACCAATGCTTACTGCTTACATGCTTACATGCTTACATGCTTACATGCTTACAT 2469  
QY 252 HistThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
DB 2470 CATACCAAGTATATCTGAGTCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2529  
QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHis 291

DB 2530 TATCTCAGAGACACACAGCATTAAGTCTCTGATCTCGAACAATTATGACTGCAT 2589  
QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuValIle 310  
DB 2590 CTTTGCATCACTCTCAAGAAACCTTGATGATCAGCTATTTGCTCTAATAAGTATGA 2646  
RESULT 2  
US-10-170-789-39  
Sequence 39, Application US/10170789  
Publication No. US20030180930A1  
GENERAL INFORMATION:  
APPLICANT: Rachel E. Meyers  
APPLICANT: Olandt, Peter J.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
APPLICANT: Welch, Nadine  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
FILE REFERENCE: 10448-191001  
CURRENT APPLICATION NUMBER: US/10/170,789  
CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797,039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882,166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212,078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934,406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226,740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861,801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19  
PRIOR APPLICATION NUMBER: US 09/801,267  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07138  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,454  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/829,671  
PRIOR FILING DATE: 2001-04-10  
PRIOR APPLICATION NUMBER: PCT/US01/40483  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: US 60/197,508  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/961,721  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: PCT/US01/29904  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: US 60/235,023  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US 10/045,367  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: US 60/246,561  
PRIOR FILING DATE: 2000-11-07  
PRIOR APPLICATION NUMBER: US 09/801,275  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07074  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,420

PRIOR FILING DATE: 2000-03-07  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 39  
LENGTH: 2649  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-170-789-39

## Alignment Scores:

Pred. No.:	1,47e-209	Length:	2649
Score:	1645.50	Matches:	310
Percent Similarity:	86.35%	Conservative:	0
Best Local Similarity:	86.35%	Mismatches:	0
Query Match:	97.95%	Indels:	49
DB:	12	Gaps:	1

US-10-070-464-3 (1-310) x US-10-170-789-39 (1-2649)

```
QY 1 PheGluGlyThrIysAspSerProLeuGluHisIleuTyValSerTyValAsn 20
DB 1570 TTGGAAGGACCAAGAGCTCCCTTTAGAGCATCAGTACAGTACAGTAAAT 1629
QY 21 ProGlyValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40
DB 1630 CCGAGAGAGGTGACCAAGGCTGACGACCGGTGCTACTCATCTTCTGCTACAGCAG 1689
QY 41 HisCysAspPhePheIleSerIysTyrSerAsnGlnIysAsnProHisCysValSerLeu 60
DB 1690 CACTGTGACTCTTTAAGATAGTATAGTAAACAGAAATCCACACTGTGTCCCTT 1749
QY 61 TyrLeuLeuSerSerProGluAspAspProThrCysIleThrIysGluPheTPAlaThr 80
DB 1750 TACAGGTATCAAGTCTCTGAAGATGACCAACTTGCAAAACAAAGAAATTTGGCCACC 1809
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGln 100
DB 1810 ATTTTGATTCACAGAGTCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 1869
QY 101 SerThrThcGlyPheThrIleuTyrgLysMetLeuTyrIysProHisAspLeuGlnProGly 120
DB 1870 AGTACTGCTGATTTACATTTGATGAGATGCTCTACAGAGCTCATGATCTACAGCTGGA 1929
QY 121 LysIleTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
DB 1930 AAGAAATATCTCTAGTGTCTCATATATGATGCTCTCAGAGTCAAGTGTGAATAT 1989
QY 134 ----- 134
DB 1990 CGGTTAAAGAGTCAAGTATTTCCGCTGAATACCTAGCTCTAGGTATGTGTT 2049
QY 134 ----- 134
DB 2050 GTAGTATGACAAACAGGGATCTGTACCGAGGGCTTAAATTGAAGCGGCTTTAA 2109
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
DB 2110 TATAAATATGCTCAATATGAATTTGACATCAGATCAGTGAAGATCTCAATCTCACTCT 2169
QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyTyr 171
DB 2170 CGATATGATTTCACTTACCTAGATCGTGTGGCATCCACCGGTGCTGCTTATGAGAGTAC 2229
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB 2230 CTCTCCCTGATGCAATTAATGACAGAGTCAAGATATCTTCAGGGGTGCTATGCTGGGGCC 2289
QY 192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
DB 2290 CCAAGTCACTGTGATCTCTCTATGATACAGGATACAGGAACTTATATAGGTCAACCT 2349
QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
```

## RESULT 3

```
US-09-976-674-2
; Sequence 2, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2671
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-2
```

## Alignment Scores:

Pred. No.:	1,49e-209	Length:	2671
Score:	1645.50	Matches:	310
Percent Similarity:	86.35%	Conservative:	0
Best Local Similarity:	86.35%	Mismatches:	0
Query Match:	97.95%	Indels:	49
DB:	9	Gaps:	1

US-10-070-464-3 (1-310) x US-09-976-674-2 (1-2671)

```
QY 1 PheGluGlyThrIysAspSerProLeuGluHisIleuTyValSerTyValAsn 20
DB 1577 TTGGAAGGACCAAGAGCTCCCTTTAGAGCATCAGTACAGTACAGTAAAT 1636
QY 21 ProGlyValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40
DB 1637 CCGAGAGGTGACCAAGGCTGACGACCGGTGCTACTCAATCTTGTGCTACAGCAG 1696
QY 41 HisCysAspPhePheIleSerIysTyrSerAsnGlnIysAsnProHisCysValSerLeu 60
DB 1697 CACTGTGACTCTTTAAGTATAGTATAGTAAACCAAGAAATCCACACTGTGTCTCTT 1756
QY 61 TyrLeuLeuSerSerProGluAspAspProThrCysIleThrIysGluPheTPAlaThr 80
DB 1757 TACAACTTATCAAGTCTTAAGATGATAGTAAACCAAGAAATTTTGGGGCCACC 1816
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGln 100
DB 1817 ATTTTGATTCACAGAGTCTCTCTGCTGACTATCTCTCCCAAAATTTCTTTTGA 1876
QY 101 SerThrThcGlyPheThrIleuTyrgLysMetLeuTyrIysProHisAspLeuGlnProGly 120
```

Db 1877 AGTACTACTGATTTACATTTGATGGATGCTCTACAGCCTCATGATCTACAGCCTGGA 1936  
Qy 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134  
Db 1937 AAGAAATATCTTACTGCTGCTTATATATGCTGCTTCTCAGGTCAGCTTGGAATAT 1996  
Qy 134 ----- 134  
Db 1997 CGATTTAAGAGATCAAGTATTTCCGCTGAATACCTTACCTCTAGGTTATGTGTT 2056  
Qy 134 ----- 134  
Db 2057 GTAGTAGTAGACAACAGGGATCTCTGACCCAGGCGCTTAATTTGAAGCGCCTTTAA 2116  
Qy 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151  
Db 2117 TATATAATGGGTGCAATAGAAATGAGATCAGTGGTGAAGAATCCCAATCTTACGTTCT 2176  
Qy 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 171  
Db 2177 CGATATGATTTTCATGATGATTCGATGATCGTGGGATCCAGCGCTGGTCTTATGGAGATAC 2236  
Qy 172 LeuSerLeuMetIleLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
Db 2237 CTCTCCCTGATGGCATTAATGACAGAGATCAGATATCTTCAGGCTTCTTATGCTGGGACC 2296  
Qy 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
Db 2297 CCAAGTCACTCTGGATGATCTTATGATACAGGATACAGGAACTTATAGGGTCAACCT 2356  
Qy 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231  
Db 2357 GACCAGATGATAACAGGGCTATTTACTTATGATCTGTGGCCATGCAAGAGAAAGTTCCCC 2416  
Qy 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
Db 2417 TCTGAACCAATGCTTACTGCTCTTACATGCTTCTTCGATGAGATGCTCATTTTGCA 2476  
Qy 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
Db 2477 CATACAGATATATATGATGATTTTATGATGAGGCTGGAAGCCATATGATTTACAGATC 2536  
Qy 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHis 291  
Db 2537 TATCCCTCAGAGACACACATTAAGATTCCTGAATCGGAGAACATTTATGAACGTCAT 2596  
Qy 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 310  
Db 2597 CTTTGCACCTACCTTCAAGAAACCTTGATCAGTATTTGCTGCTTAAAGTGATA 2653

RESULT 4  
US-10-170-789-37  
; Sequence 37, Application US/10170789  
; Publication No. US20030180930A1  
; GENERAL INFORMATION:  
; APPLICANT: Rachel E. Meyers  
; APPLICANT: Olandt, Peter J.  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Curtis, Rory A. J.  
; APPLICANT: Williamson, Mark  
; APPLICANT: Weich, Nadine  
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
; TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF  
; FILE REFERENCE: 10448-191001  
; CURRENT APPLICATION NUMBER: US/10/170,789  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 09/797,039  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: PCT/US01/06525  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/186,061  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 09/882,166

; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: PCT/US01/19269  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/212,078  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 09/934,406  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: PCT/US01/26052  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: US 60/226,740  
; PRIOR FILING DATE: 2000-08-21  
; PRIOR APPLICATION NUMBER: US 09/861,801  
; PRIOR FILING DATE: 2001-05-21  
; PRIOR APPLICATION NUMBER: PCT/US01/16549  
; PRIOR FILING DATE: 2001-05-21  
; PRIOR APPLICATION NUMBER: US 60/205,508  
; PRIOR FILING DATE: 2000-05-19  
; PRIOR APPLICATION NUMBER: US 09/801,267  
; PRIOR FILING DATE: 2001-03-06  
; PRIOR APPLICATION NUMBER: PCT/US01/07138  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: US 60/187,454  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: US 09/829,671  
; PRIOR FILING DATE: 2001-04-10  
; PRIOR APPLICATION NUMBER: PCT/US01/40483  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: US 60/197,508  
; PRIOR FILING DATE: 2000-04-18  
; PRIOR APPLICATION NUMBER: US 09/961,721  
; PRIOR FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: PCT/US01/29904  
; PRIOR FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: US 60/235,023  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US 10/045,367  
; PRIOR FILING DATE: 2001-11-07  
; PRIOR APPLICATION NUMBER: US 60/246,561  
; PRIOR FILING DATE: 2000-11-07  
; PRIOR APPLICATION NUMBER: US 09/801,275  
; PRIOR FILING DATE: 2001-03-06  
; PRIOR APPLICATION NUMBER: PCT/US01/07074  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: US 60/187,420  
; PRIOR FILING DATE: 2000-03-07  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 37  
; LENGTH: 3143  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (229)...(2874)  
US-10-170-789-37

Alignment Scores:  
Pred. No.: 1,94e-209 Length: 3143  
Score: 1645.50 Matches: 310  
Percent Similarity: 86.35% Conservative: 0  
Best Local Similarity: 86.35% Mismatches: 49  
Query Match: 97.95% Indels: 1  
DB: 12 Gaps: 1

US-10-070-464-3 (1-310) x US-10-170-789-37 (1-3143)

Qy 1 PheGluGlyThrIleAspSerProLeuGluHisIleLeuTyrValIleSerTyrValAsn 20  
Db 1798 TTTGAAGGACCCAAAGATCTCCCTTTAGAGATCAGTCTGATGATGATGATGAT 1857  
Qy 21 ProGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln 40  
Db 1858 CCTGAGAGAGTGACAAAGGCTGATGACCGGTCTACTCATCTTGTGTCATCAGTCAG 1917

Qy 41 HisCysAspPhePhe1LeSerLyTySerAenGlnLysAsnProHisCysValSerLeu 60  
Db 1918 CACTGTGACTTCTTATAGTAGTAGTAAACGAAAGAAATCCACACTGTGTCTCTT 1977  
Qy 61 TyTyLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 80  
Db 1978 TACAAGCTATCAAGCTCGAAGATGACCCCACTTCCAAACAAAGAAATTTGGGCCACC 2037  
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
Db 2038 ATTGTGGATCAAGAGGCTCTCTCTTCTGACTACTCTCCAGAAATTTCTCTTTTGA 2097  
Qy 101 SerThrThrGlyPheThrLeuTyrgLysLeuTyrgLysProHisAspLeuGlnProGly 120  
Db 2098 AGTACTACTGATTTTACATTTGATGAGATGCTCTCAAGCCTCAGATCTACAGCCTGGA 2157  
Qy 121 LysLysTygProThrValLeuPheIleTyrgLysProGln----- 134  
Db 2158 AAGAATATCTCTACTGTCTGTTCATATATGATGCTCTCAGGTGCACTGTGTAATAT 2217  
Qy 134 ----- 134  
Db 2218 CGGTTTAAAGATCAAGTATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTGTT 2277  
Qy 134 ----- 134  
Db 2278 GTAGTAGTAGACAAAGGGGATCTCTGACAGAGGGCTTAATTGAAGGGCCTTTAA 2337  
Qy 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151  
Db 1338 TATATAATGGGTCAAAATGAATGACATCAGATCAGGAGGAGCTCCAAATATCTAGCTCT 2397  
Qy 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrgLysTyg 171  
Db 2398 CGATATGATTTTACTTACTTATGATGCTGTGGCATCCAGGCTGTCTCTATGGAGAAAC 2457  
Qy 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
Db 2458 CTCCTCCGATGGCATTAATGACAGAGGTCAAGTATCTTCAAGGTTGCTATTTGGGGCC 2517  
Qy 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
Db 2518 CCACTACTCTGTGATCTCTTATGATACAGATACAGGAAACGTTATATGGGTCAACCT 2577  
Qy 212 AspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231  
Db 2578 GACCAGAAATGACAGGGCTATTTACTTAGGATCTGTGGCATGCAAGCAAGAAAGTTCCCC 2637  
Qy 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
Db 2638 TCTGAAACCAATTCGTTTACTGCTCTTACATGTTTCCGATGAGAAATGCTCATTTTCCA 2697  
Qy 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
Db 2698 CATACCAATATATATGATGATTTTACTGAGGGCTGGAAGCCATATGATTTACAGATC 2757  
Qy 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHis 291  
Db 2758 TATCTCAAGGAGAACACAGCATTAAGTTCCTTAATCTGGGAGCAATTTAAGAACTGCA 2817  
Qy 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuValIle 310  
Db 2818 CTTTGTGACTTACTTCAAGAAACCTTGATCAGTATGCTGTCTTAAGTATATA 2874

APPLICANT: Junten, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 20  
LENGTH: 4676  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-976-674-20  
Alignment Scores:  
Pred. No.: 5,05e-209 Length: 4676  
Score: 1644.50 Matches: 308  
Percent Similarity: 99.35% Conservative: 0  
Best Local Similarity: 99.35% Mismatches: 1  
Query Match: 97.89% Indels: 2  
DB: 10 Gaps: 1  
US-10-070-464-3 (1-310) x US-09-976-674-20 (1-4676)  
Qy 1 PheGluGlyThrLysAspSerProLeuGlnHisLysLeuTyrgValSerTyrgValAsn 20  
Db 1783 TTTGAAGCACCAAGACTCCCTTTAGAGATCACTGTGACGATGATGATGATGATGAT 1842  
Qy 21 ProGlyGluValThrArgLeuThrAspArgGlyTygSerHisSerCysLeuSerGln 40  
Db 1843 CCGGAGAGGAGACAGGCTGACGACCGGTGCTACTACATTTCTGTGATCAGTCAAG 1902  
Qy 41 HisCysAspPhePhe1LeSerLyTySerAenGlnLysAsnProHisCysValSerLeu 60  
Db 1903 CACTGTGACTTCTTATAGTATAGTATAGTAAACGAAAGAAATCCACACTGTGTCTCT 1962  
Qy 61 TyTyLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 80  
Db 1963 TACAAGCTATCAAGCTCTGAAAGATGACCAACTGCAAAACAAAGAAATTTGGGCCACC 2022  
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
Db 2023 ATTTGGATTCAGAGAGGCTCTCTCTGACATATCTCTCCAGAAATTTCTCTTTGAA 2082  
Qy 101 SerThrThrGlyPheThrLeuTyrgLysLeuTyrgLysProHisAspLeuGlnProGly 120  
Db 2083 AGTACTACTGATTTTACATTTGATGAGATGCTCTTCAAGCCTCATGATCTACAGCCTGA 2142  
Qy 121 LysLysTygProThrValLeuPheIleTyrgLysProGlnGlyGlnIleGluIleAsp 140  
Db 2143 AAGAATATCTCTACTGTGCTGTTCATATATGTGTGTGCTGCTGCTGCTGCTGCTGCT 2198  
Qy 141 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrgAspPheIleAspLeuAspArg 160  
Db 2199 GATAGGTGAGAGAGATCCCAATATCTAGCTTCTGATATGATTTTCACTTACATGATCT 2258  
Qy 161 ValGlyIleHisGlyTyrPheSerTyrgLysTyrgLysLeuSerLeuMetAlaLeuMetGlnArg 180  
Db 2259 GTGGGATTCACAGGCTGCTCTTATGAGAGATACCTCTCCGATGGCATTAATGACAGG 2318  
Qy 181 SerAspIlePheAspValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAsp 200  
Db 2319 TCGATATCTTCAAGGTTGCTATTGCTGGGGCCCCACATCTGTGTGATCTTATGAT 2378  
Qy 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeu 220  
Db 2379 ACAAGATACACAGGAACTTATATGGGTCACTTACACCAAGATGAACAGGCTATTTACTTA 2438  
Qy 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 240  
Db 2439 GATCTGTGGCATGACAGCAAGAAAGTCCCTCTGAAACCAATCGTTTACTGCTCTTA 2498

RESULT 5  
US-09-976-674-20  
Sequence 20, Application US/0976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Akisanya, Karen  
APPLICANT: Riviere, Pierre

QY 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisSerSerLeuLeuSerPheLeu 260  
DB 2499 CATGTTTCCTCGATGAGAAATGTCATTTTGACATACAGATATTAAGTATTAAGTATTTT 2558  
QY 261 ValAlaGlaGlyLysProThrAspLeuGlnIleTyrProGlnGlnArgHisSerIleArg 280  
DB 2559 GTGAGGGCTGGAGAAAGCCATATGATTTTACAGATCTATCTCAGAGAAAGACACGCAATAGA 2618  
QY 281 ValProGluSerGlyGlnHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnLeu 300  
DB 2619 GTTCTGTAATCGGAGAACATTTATGAACTGCATCTTTTGCACTACCTTCAAGAAAACCTT 2678  
QY 301 GlySerArgIleAlaAlaLeuLysValIle 310  
DB 2679 GCATCAGTATTGCTGCTCTTAAGATGATA 2708  
RESULT 6  
US-09-976-674-14  
; Sequence 14, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Oi, Steve  
; APPLICANT: Akinsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 4309  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-976-674-14  
Alignment Scores:  
Pred. No.: 1,566-206 Length: 4309  
Score: 1625.50 Matches: 310  
Percent Similarity: 85.87% Conservative: 0  
Best Local Similarity: 85.87% Mismatches: 0  
Query Match: 96.76% Indels: 51  
DB: 10 Gaps: 1  
US-10-070-464-3 (1-310) x US-09-976-674-14 (1-4309)  
QY 1 PheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSerTyrValAsn 20  
DB 1263 TTTGAAGGCACCAAGACTCCCTTTAGAGCATCACTTACGTAGTACGTTAAAT 1322  
QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysValSerGln 40  
DB 1323 CTTGGAGAGGTACCAAGGCTGACTGACCGTGGTACTCATCTTCTGCACTGACTG 1382  
QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60  
DB 1383 CACTGTGACTCTTTATATAGTAAGTATTAACCAAGAAATCCACACTGTGTGCCCTT 1442  
QY 61 TyrIleValSerSerProGluLysAspProThrCysIleThrLysGluPheThrAlaThr 80  
DB 1443 TACAGACTATCAAGCTCTGAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCGCAC 1502  
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGlnIlePheSerPheGlu 100  
DB 1503 ATTTTGATTCACGAGTCTCTCTCTGACTATACCTCCCAAGAAATTTCTCTTTGA 1562  
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
DB 1563 AGTACTACTGATTAATTAATTTATGAGATGCTTCAAGGCTCATGATCTACAGCCTGGA 1622

QY 121 LysIleTyrProThrValLeuPheIleTyrGlyGly--ProGln----- 134  
DB 1623 AAGAAATATCTACTGCTGTCTCATATATGTTGGTCTCTCAGGTGCAATTGTGAATA 1682  
QY 134 ----- 134  
DB 1683 ATCGGTTAAAGAGCTAAGTATTTCCGCTTGAAATCCCTAGCTCTTAGTTATGTGG 1742  
QY 134 ----- 134  
DB 1743 TTGTAGTATAGACAACAGGGATTCCTGCACCGAGGCTTAATTTGAAGCGCCTTTA 1802  
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 151  
DB 1803 AATATTAATATGCTCAATATGAATAATTTGACGATCAGGTGGAAGACTCCAAATATCTAGCTT 1862  
QY 151 GATAGTThrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyT 171  
DB 1863 CTGATATGATTTTCACTTACTTATGATCGTGGGATCCAGGCTGCTGCTTATGAGAGAT 1922  
QY 171 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 191  
DB 1923 ACCTTCCCTGATGCGCATTAATGCAAGCTCAGATATCTTCAGGCTTCTATGCTGGGG 1982  
QY 191 IAProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisP 211  
DB 1983 CCCAGTCACTGTGGATCTTCTATGATACAGATACACGGAAGTTATATGGGCTCAC 2042  
QY 211 RoAspGluAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnIleGlnLysPheP 231  
DB 2043 CTGACCAATAAGAACAGGCTTATTACTTAAGATCTGTGGCCATGACAGCAAGAAAGTTCC 2102  
QY 231 RoSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251  
DB 2103 CCTCGAACCAATGCTTACTGCTCTTACATGCTTCTCGATAGAAATGCCATTTTG 2162  
QY 251 IAHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 271  
DB 2163 CACATACCAATATATATCTGAGTTTATGAGAGGCTGGAAGCCATATGATTTTACAGA 2222  
QY 271 IeTyrProGlnGlnArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeu 291  
DB 2223 TCTATCTCAGAGAGACACAGCATTAAGTCTCGAATCGGAGAACATTTATGAATGCC 2282  
QY 291 IeLeuLeuHisTyrLeuGlnGlnLeuGlySerArgIleAlaAlaLeuLysValIle 310  
DB 2283 ATCTTTGCATCACTCTCAAGAAAACCTTGATCACTGATTGCTGCTCAAAAGTGATA 2341  
RESULT 7  
US-09-976-674-12  
; Sequence 12, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Oi, Steve  
; APPLICANT: Akinsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 4829  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-976-674-12  
Alignment Scores:  
Pred. No.: 1,886-206 Length: 4829

Score: 1625.50 Matches: 310  
 Percent Similarity: 85.87% Conservative: 0  
 Best Local Similarity: 85.87% Mismatches: 0  
 Query Match: 96.76% Indels: 51  
 DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x US-09-976-674-12 (1-4829)

QY 1 PheGluGlyThrIleuSerProLeuGluHisIleuTyValIleSerTyValAsn 20  
 DB 1783 TTTGAAGGACCAAGACCTCCCTTAAAGACATCACTGACGATGACGTAAT 1842  
 QY 21 ProGlyValuValThrIleuSerProLeuGluHisIleuTyValIleSerTyVal 40  
 DB 1843 CCGAGAGAGGTGACCAAGGCTGACGACCGGCTACTACCAATCTTCTGATCAGCAG 1902  
 QY 41 HisCysAspPhePheIleSerTyValSerAsnGluAsnProHisCysValSerLeu 60  
 DB 1903 CACTGTGACCTCTTTATTAAGTATGATGATGATGATGATGATGATGATGATGAT 1962  
 QY 61 TyIleuSerProGluAspPheProThrCysIleuThrIleuSerTyValIleSer 80  
 DB 1963 TACAGATATCAAGTCTGAGAGGACCACTGCAAAACAAAGAAATTTGGGCACC 2022  
 QY 81 IleuAspSerIleuGlyProLeuProAspTyThrProGluIlePheSerPheGlu 100  
 DB 2023 ATTTGGATTACAGAGTCTCTCTTCTGATATCTCTGCAAAATTTCTCTTTGAA 2082  
 QY 101 SerThrThrGlyPheThrIleuTyValIleuSerTyValIleuSerTyValIle 120  
 DB 2083 AGTACTACTGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2142  
 QY 121 LysIleuTyProThrValIleuPheIleTyGlyGly--ProGln----- 134  
 DB 2143 AAGAAATATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2202  
 QY 134 ----- 134  
 DB 2203 ATCGGTTTAAAGAGTCAATATTTCCGCTTGAATACCTGACCTCTGATGATGATG 2262  
 QY 134 ----- 134  
 DB 2263 TTGTAGTATGACCAAGGAGATCTCTGACCGAGGCTTAAATTTGAAGGCGCTTTA 2322  
 QY 135 -----GlyGlnIleGluIleAspAspGlnValIleGluIleuGlnTyLeuAs 151  
 DB 2323 AATATATAATGGGTCAATATGAATGATGATGATGATGATGATGATGATGATGATG 2382  
 QY 151 eArGlyAspPheIleAspLeuAspArgValGlyIleHisGlyTTPSerTyGlyYT 171  
 DB 2383 CTGATATATGATTTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 2442  
 QY 171 TyIleuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 191  
 DB 2443 ACCCTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2502  
 QY 191 IaProValThrIleuTyPheTyAspThrGlyTyThrGluIleuGlyMetGlnHisP 211  
 DB 2503 CCCAGTCACTGCTGATTTCTATGATGATGATGATGATGATGATGATGATGATGATG 2562  
 QY 211 roAspGlnAsnGluGlnTyTyIleuGlySerValAlaMetGlnAlaGluTyPheP 231  
 DB 2563 CTGACCAAGATGAAGGCTATTAATCTAGATCTGAGGCTGAGGCAAGCAAGAAAGTTC 2622  
 QY 231 roSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 251  
 DB 2623 CCTCTAAACCAATCGTTTACTCTCTTACATGATGATGATGATGATGATGATGATG 2682  
 QY 251 IaHisThrSerIleuLeuSerPheLeuValArgAlaGlyLysProTyAspLeuGln 271  
 DB 2683 CACATACCACTATATTAATGATTTTATGATGATGATGATGATGATGATGATGATG 2742  
 QY 271 IaTyProGluGlnArgHisSerIleArgValProGluSerGlyGluHisTyGluLeu 291

DB 2743 TCTATCTGAGAGACACAGCATTAAGTTCTGTAATGGGAGAACTATTAAGTCTG 2802  
 QY 291 IseuLeuHisTyIleuGlnGluAsnLeuGlySerArgIleAlaIleuValIle 310  
 DB 2803 ATCTTTGCACTACTCTTCAAGAAACCTTGATCAGATGATGCTGCTTAAAGTATA 2861

# RESULT 8

US-09-976-674-22  
 ; Sequence 22, Application US/09976674  
 ; Patent No. US20020115843A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Qi, Steve  
 ; APPLICANT: Akimanya, Karen  
 ; APPLICANT: Riviere, Pierre  
 ; APPLICANT: Juntien, Jean-Louis  
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
 ; FILE REFERENCE: 70669  
 ; CURRENT APPLICATION NUMBER: US/09/976,674  
 ; PRIOR FILING DATE: 2001-10-12  
 ; PRIOR FILING DATE: 2000-10-12  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 22  
 ; LENGTH: 4685  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-976-674-22

## Alignment Scores:

Pred. No.: 1,176-170 Length: 4685  
 Score: 1358.50 Matches: 271  
 Percent Similarity: 88.61% Conservative: 9  
 Best Local Similarity: 85.76% Mismatches: 27  
 Query Match: 80.86% Indels: 9  
 DB: 10 Gaps: 5

US-10-070-464-3 (1-310) x US-09-976-674-22 (1-4685)

QY 1 PheGluGlyThrIleuSerProLeuGluHisIleuTyValIleSerTyValAsn 20  
 DB 1783 TTTGAAGGACCAAGACCTCCCTTAAAGACATCACTGACGATGACGTAAT 1842  
 QY 21 ProGlyValuValThrIleuSerProLeuGluHisIleuTyValIleSerTyVal 40  
 DB 1843 CCGAGAGAGGTGACCAAGGCTGACGACCGGCTACTACCAATCTTCTGATCAGCAG 1902  
 QY 41 HisCysAspPhePheIleSerTyValSerAsnGluAsnProHisCysValSerLeu 60  
 DB 1903 CACTGTGACCTCTTTATTAAGTATGATGATGATGATGATGATGATGATGATGAT 1962  
 QY 61 TyIleuSerProGluAspPheProThrCysIleuThrIleuSerTyValIleSer 80  
 DB 1963 TACAGATATCAAGTCTGAGAGGACCACTGCAAAACAAAGAAATTTGGGCACC 2022  
 QY 81 IleuAsp-----SerIleuGlyProLeuProAspTyThr-----ProGluIleP 97  
 DB 2023 ATTTGGATTACAGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 2082  
 QY 97 eSerPheGluSerThrThrGlyPheThrIleuTyValIleuSerTyValIleSer 116  
 DB 2083 TCGGCTTGAATACCTTACGCTCTCTATGATGATGATGATGATGATGATGATGATG 2142  
 QY 117 -LeuGlnPro-GlyIleuTyTyProThrValIleuPheIleTyGlyGlyProGlnGly 136  
 DB 2143 CCGTCAACCAAGGCTTAAATTTGAAGGCGCC---TTTAAATAT-----AAAATGGG 2193  
 QY 136 ImlIleuIleAspAspGlnValIleGlyIleuGlnTyIleuAlaSerArgTyAspPhe 156  
 DB 2194 AATATATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2253  
 QY 156 IaAspLeuAspArgValIleHisGlyTTPSerTyGlyGlyTyIleuSerLeuMetAl 176



Db 2254 TTGACTTAGATCGTGTGGCATCCACGGCTGTCCTATGAGAGTAACCTCCCTGATGG 2313  
Qy 176 laeuMetGlnArgSerAspIlePheArgValAlaIleAlaProValThrLeuT 196  
Db 2314 CATTAAAGCAGAGTCAGATATCTTCAGGGTGTATATGCGGGCCCCAGTCACCTGT 2373  
Qy 196 rPilePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlu 216  
Db 2374 GGATCTTCTATGATACAGATACACGGACGTTATATGGGTCAACCCGACACAGATGAC 2433  
Qy 216 lngIlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsn 236  
Db 2434 AGGCTATTAAGTACTGATCTGTGGCCATGCACAGCAAAAGTCCCTCTGAACCAATC 2493  
Qy 236 rglLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIle 256  
Db 2494 GTTTACTGCTCTTACATAGTGTCTCTGATGAGATGCAATTTTGACATACACAGATAT 2553  
Qy 256 euleuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGlu 276  
Db 2554 TACTGAGTTTATTAAGTAGGGCTGGAAAGCCATATGATTTACAGATCTATCTCAGAGA 2613  
Qy 276 rglHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyr 296  
Db 2614 GACACAGCATAGAGGTTCCGAAATCGGAGAACATATGAACTGCATCTTTGACATACC 2673  
Qy 296 eugGlnGluAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310  
Db 2674 TTCAAGAAACCTTGATCAAGTATCTGCTCTTAAGATGATA 2717  
RESULT 9  
US-09-976-674-8 ; Sequence 8, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akinsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 4523  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-976-674-8  
Alignment Scores:  
Pred. No.: 8.23e-170 Length: 4523  
Score: 1352.00 Matches: 257  
Percent Similarity: 82.90% Conservative: 0  
Best Local Similarity: 82.90% Mismatches: 1  
Query Match: 80.48% Indels: 53  
DB: 10 Gaps: 1  
US-10-070-464-3 (1-310) x US-09-976-674-8 (1-4523)  
Qy 1 PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSerTyrValAsn 20  
Db 1783 TTTGAAGGCAACCAAGACTCCCTTTAGAGCATACCTGTACGTAAGTACAGTAAAT 1842  
Qy 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysValIleSerGln 40  
Db 1843 CCTGAGAGGCTGACAAAGCTGACGACCGTGGCTACTACATCTTCTGCTGACATGACG 1902  
Qy 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60

Db 1903 CACTGAGACTTCTTTAAGTAAAGTATAGTAACCAAGAAATCACACTGTGTCTCCTT 1962  
Qy 64 TyrLysLeuSerSerProGluAspAspProThrCysLysThrTyrGluGluPheTyrAlaThr 80  
Db 1963 TACAACTATCAAGTCTCTGAAGATGACCCAACTTGCAAAACAAAGAAATTTTGAGCCACC 2022  
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
Db 2023 ATTTTGATTCACAGAGTCTCTCTCTGACTATACCTCTCCAAATTTTCTCTTTGAA 2082  
Qy 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
Db 2083 AGTACTACTGATTTACATGTATGAGGATCTCTACAAAGCTCATGATCTACAGCTGGA 2142  
Qy 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnGlnIleGluIleAsp 140  
Db 2143 AAGAAATATCTTACTGCTGTTCATATATGATGTTGTT----- 2178  
Qy 141 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160  
Db 2178 ----- 2178  
Qy 161 ValGlyIleHisGlyTyrPserTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180  
Db 2178 ----- 2178  
Qy 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrPheTyrAsp 200  
Db 2179 -----CGGTT-GCTATGCTGGGGCCCCAGTCCTGTGGATCTTCTATGAT 2225  
Qy 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeu 220  
Db 2226 ACAGATATACCGGAAGCTTATATGGGTACCCGACCAAGATGAACAGGCTATTACTTA 2285  
Qy 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 240  
Db 2286 GGATCTGTGGCCATGCAGACGAAAGTTCCCTCTGAACCAATCGTTTACGCTCTTA 2345  
Qy 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 260  
Db 2346 CATGGTTCTCTGATGAGATGCAATGTCATTTTGACATACAGATATTAAGTGAATTTTA 2405  
Qy 261 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 280  
Db 2406 GTGAGGCTGGAAGCCATATGATTTACAGATCTATCTCAGAGACACAGCATAGA 2465  
Qy 281 ValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 300  
Db 2466 GTTCCTGAATCGGAGAACATATGAACTGCATCTTTGACATACCTTCAAGAAACCTT 2525  
Qy 301 GlySerArgIleAlaIleAlaLeuLysValIle 310  
Db 2526 GGATCAAGTATCTGCTCTTAAGATGATA 2555  
RESULT 10  
US-09-976-674-4 ; Sequence 4, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akinsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4

LENGTH: 2617  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-976-674-4

## Alignment Scores:

Alignment No.:	1,59e-135	Length:	2617
Score:	1093.50	Matches:	200
Percent Similarity:	69.34%	Conservative:	42
Best Local Similarity:	57.31%	Mismatches:	58
Query Match:	65.09%	Indels:	49
	10	Gaps:	1

US-10-070-464-3 (1-310) x US-09-976-674-4 (1-2617)

```

QY 1 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValaen 20
   |||.....|
Db 1553 TTCACGGGACCAAGACAGCGCGCTGAGACACACCTTACGTGCTAGTATGAGGG 1612
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
   |||.....|
Db 1613 GCGGGGAGATGCTACCGCTCCACACCGCGCTTCCCATAGCTGCTCCATGAGCAG 1672
QY 41 HisCysAspPhePheIleSerIlyTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
   |||.....|
Db 1673 AACTTCGACATGCTGCTGACGACCTACAGCAGCGTGAAGCAGCCCGCTGCTGACAGTC 1732
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrPalaThr 80
   |||.....|
Db 1733 TACAAAGCTAGGCGGCGCCGACACAGCCCTGACACACAGCCCGCTTCTGGGCTAGC 1792
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
   |||.....|
Db 1793 ATGATGAGAGGAGCGACGCGCGCGGATATGTTCTCCAGAGATCTTCATTTCCAC 1852
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
   |||.....|
Db 1853 ACGCGCTCGAGATGCTGCGCTTACGCGCATGATCTCAAGCCCGCAGCTTGACAGGAG 1912
QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
   |||.....|
Db 1913 AAGAAGCACCCACGCTCTTTGTATATGAGGCGCCAGGTGACGCTGTGATATAC 1972
QY 134 ----- 134
Db 1973 TCCTTCAAGGATCAAGTACTTGGCGCTCAACACACTGCGCTCCCTGGGCTACGCGCTG 2032
QY 134 ----- 134
Db 2033 GTTGTGATTGACGGGAGGCGCTCTGTACGAGGAGGCTTCGTTGCAAGGCGCGCTGA 2092
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
   |||.....|
Db 2093 AACCAATGGGCGGAGTGAAGTCAAGAGCAGAGTGAAGGCGCTGAGTCTTGCGCGAG 2152
QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheTyrGlyGlyTyr 171
   |||.....|
Db 2153 AAGTATGCTTCATGACCTGAGCGAGTGCATTCATGAGTGCATGAGCGGCGCTTC 2212
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
   |||.....|
Db 2213 CTCTGCTCATGCGGCTAATCCACAAAGCCCGAGGTGTTTCAAGGTGGCCATCCGCGG 2272
QY 192 ProValThrLeuThrIlePheTyrAspThrArgIlyTyrThrGluArgTyrMetGlyHisPro 211
   |||.....|
Db 2273 CCGGTCACCGTGTGATGCGCTTACAGACAGGCTTACAGCGCTTACAGCGCTTCCT 2332
QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
   |||.....|
Db 2333 GAGAAACACGACGAGCGGTATGAGCGGCTTCGTTGCGCTGACGTGAGAGAGCTGCC 2392
QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
   |||.....|

```

```

Db 2393 AATGAGCCCAACCGCTTGCTTATCTCCACGCGCTTCTGAGCAAAACGTGACCTTTTC 2452
QY 252 HisThrSerIleLeuLeuSerPheLeuValAlaArgAlaGlyLysProTyrAspLeuGlnIle 271
   |||.....|
Db 2453 CACACAACTTCTGCTCTCCACTGATCGACAGGAGAACTTTACAGCTCCGACATC 2512
QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlyHisTyrGluLeuHis 291
   |||.....|
Db 2513 TACCCCAACGAGACAGCATATTCGCTGCGCCCGAGTCCGGGAGAGCATATGAGTCAG 2572
QY 292 LeuLeuHisTyrLeuGlnGluAsnLeu 300
   |||.....|
Db 2573 TTGCTGACATTTCTACAGGAATATCCTC 2599

RESULT 11
US-09-976-674-28
; Sequence Db, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akisanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Juntzen, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 4219
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-976-674-28

Alignment Scores:
Pred. No.: 3,47e-135 Length: 4219
Score: 1093.50 Matches: 200
Percent Similarity: 69.34% Conservative: 42
Best Local Similarity: 57.31% Mismatches: 58
Query Match: 65.09% Indels: 49
DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x US-09-976-674-28 (1-4219)
QY 1 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValaen 20
   |||.....|
Db 1909 TTCAGGGGACCAAGACAGCGCGCTGAGACACACCTTACGTGCTAGTATGAGGG 1968
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
   |||.....|
Db 1969 GCGCGGAGATGCTGAGCGCTTCAACAGCGCGCGCTTCCCATAGCTGCTCATGAGCAG 2028
QY 41 HisCysAspPhePheIleSerIlyTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
   |||.....|
Db 2029 AACTTCGACATGTTGCTGACGACCTACAGAGCGTGAAGCAGCCCGCTGCGTGAAGTC 2088
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrPalaThr 80
   |||.....|
Db 2089 TACAAAGCTAGGCGGCGCCGACAGACAGCCCTGACACAAAGCAGCCCGCTTCTGGGCTAGC 2148
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
   |||.....|
Db 2149 ATGATGAGAGGAGCGCGCGCTGAGTATGTTCTCCAGAGATCTTCATTTCCAC 2208
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
   |||.....|
Db 2209 ACGCGCTGAGATGTCGCGCTTACGCGCATGATCTCAAGAGCCCGCAGCGCTTGACGAGG 2268
QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
   |||.....|

```

Db 2269 AAGAACACCCACCGTCTCTTTGATATGAGAGCCCCCAGAGTGCAGCTGTGATATAC 2328  
 QY 134 ----- 134  
 Db 2329 TCCTTCAAGGATCAAGTACTGGGCTCAACACACTGGCCTCCTGGGCTACGCCGTG 2388  
 QY 134 ----- 134  
 Db 2389 GTTGTGATTGACGGCAGGGGCTCTGTACGACGAGGCTTCGGTTGGAAGGGGCCCTGAAA 2448  
 QY 135 -----GlyGlnIleGluIleAspArgIleValIleGlyTyrSerTyrGlyGlyTyr 151  
 Db 2449 AACCAATGAGGCGAGTGAATCGAGACCAAGTGGAGGCGCTGCAAGTTCGTGGCGGAG 2508  
 QY 152 ArgTyrAspPheIleAspLeuAspArgValIleGlyTyrSerTyrGlyGlyTyr 171  
 Db 2509 AAGTATGGCTTCACTGACCTGAGCCAGATGGCCATCAGCTGATGCTGATCGGGGCTTC 2568  
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
 Db 2569 CTCTGGCTCATGGGCTTAATCCACAAAGCCCCAGAGTGTTCAGAGTGGCCATCGCGGCTGC 2628  
 QY 192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
 Db 2629 CCGGTACCGGTCTGATGGCTTACGACAGAGGCTACAGAGGCTACAGAGCTTCAC 2688  
 QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluIleAspLeu 231  
 Db 2689 GAGAACACACAGACAGGCTATATGAGGCGGCTTCGTTGGCTGTGACCTGAGAAAGCTGCC 2748  
 QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
 Db 2749 AATGAGCCCAACCGCTTGTCTTATCCACAGGCTTCTGACGAAACCTGACCTTTTC 2808  
 QY 252 HistSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
 Db 2809 CACACAACTTCTCTCTCCCACTGATCCGAGCGGAGAACTTACCAAGCTCCAGATC 2868  
 QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291  
 Db 2869 TACCCCAAGAGACAGACAGTATTCGCTGCCGAGTCCGGGAGAGCACTATGAAGTCACG 2928  
 QY 292 LeuLeuHisTyrLeuGlnGluAsnLeu 300  
 Db 2929 TTGCTGCACTTCTACAGGAATACCTC 2955  
 RESULT 12  
 US-09-976-674-24  
 ; Sequence 24, Application US/09976674  
 ; Patent No. US20020115843A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Qi, Steve  
 ; APPLICANT: Akimsanya, Karen  
 ; APPLICANT: Riviere, Pierre  
 ; APPLICANT: Junien, Jean-Louis  
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV  
 ; FILE REFERENCE: 70669  
 ; CURRENT APPLICATION NUMBER: US/09/976,674  
 ; PRIOR FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: US 60/240,117  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 24  
 ; LENGTH: 4302  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-976-674-24  
 Alignment Scores:  
 Pred. No.: 3,59e-135 Length: 4302  
 Score: 1093.50 Matches: 200  
 Percent Similarity: 69.34% Conservative: 42

Best Local Similarity: 57.31% Mismatches: 58  
 Query Match: 65.09% Indels: 49  
 DB: 10 Gaps: 1  
 US-10-070-464-3 (1-310) x US-09-976-674-24 (1-4302)  
 QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValIleSerTyrValAsn 20  
 Db 1909 TTCAGGGGACCAAGGACACGCGCTGAGACCAACCTTACGTGTGACATGAGCGG 1968  
 QY 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAlaSerGln 40  
 Db 1969 GCCGCGAGATGTGACGCTTCCACACGCGCGGCTTCTCCCAATAGCTGTCCATGAGCCG 2028  
 QY 41 HisCysAspPhePheIleSerTyrTyrSerAsnGlnIleAsnProHisCysValSerLeu 60  
 Db 2029 AACTTCGACATGTGCTGACGCCATACACAGACCGTGAAGCAGCCGCTTGCAGCGTC 2088  
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 80  
 Db 2089 TACAAGCTGAGCGGCGCCGACAGACACCCCTGCACAAAGACACCCGCTTGGGCTTAC 2148  
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100  
 Db 2149 ATGATGAGGACGACGACGCTGCCCGCGATATGTCTTCCAGAGATCTTCCATTTCCAC 2208  
 QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
 Db 2209 ACGGCTCGAATGTGCGGCTTACGCGCATGATCTACAAAGCCCAAGCCTTGAGCCAGG 2268  
 QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134  
 Db 2269 AAGAACACCCACCGCTCTTGTATATGAGAGCCCCCAGAGTGCAGCTGTGATATAC 2328  
 QY 134 ----- 134  
 Db 2329 TCCTTCAAGGATCAAGTACTGGGCTCAACACACTGGCCTCCTGGGCTACGCCGTG 2388  
 QY 134 ----- 134  
 Db 2389 GTTGTGATTGACGGCAGGGGCTCTGTACGACGAGGCTTCGGTTGGAAGGGGCCCTGAAA 2448  
 QY 135 -----GlyGlnIleGluIleAspArgIleValIleGlyLeuGlnTyrLeuAlaSer 151  
 Db 2449 AACCAATGAGGCGAGTGAATCGAGACCAAGTGGAGGCGCTGCAAGTTCGTGGCGGAG 2508  
 QY 152 ArgTyrAspPheIleAspLeuAspArgValIleGlyTyrSerTyrGlyGlyTyr 171  
 Db 2509 AAGTATGGCTTCACTGACCTGAGCCAGATGGCCATCAGCTGATGCTGATCGGGGCTTC 2568  
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
 Db 2569 CTCTGGCTCATGGGCTTAATCCACAAAGCCCCAGAGTGTTCAGAGTGGCCATCGCGGCTGC 2628  
 QY 192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
 Db 2629 CCGGTACCGGTCTGATGGCTTACGACAGAGGCTACAGAGGCTTACAGAGCTTCAC 2688  
 QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluIleAspLeu 231  
 Db 2689 GAGAACACACAGACAGGCTATATGAGGCGGCTTCGTTGGCTGTGACCTGAGAAAGCTGCC 2748  
 QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251  
 Db 2749 AATGAGCCCAACCGCTTGTCTTATCCACAGGCTTCTGAGAGAAAGCTGACCTTTTC 2808  
 QY 252 HistSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271  
 Db 2809 CACACAACTTCTCTCTCCCACTGATCCGAGCGGAGAACTTACCAAGCTCCAGATC 2868  
 QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291  
 Db 2869 TACCCCAAGAGACAGACAGTATTCGCTGCCGAGTCCGGGAGAGCACTATGAAGTCACG 2928

QY	292	leuLeuH1sYrLeuGlnGluAsnLeu 300
DB	2929	TTGCTGACCTTCTACAGAAATACCTC 2955
RESULT 13		
US-09-976-674-36		
; Sequence 36, Application US/09976674		
; Patent No. US20020115843A1		
; GENERAL INFORMATION:		
; APPLICANT: Qi, Steve		
; APPLICANT: Akinsanya, Karen		
; APPLICANT: Riviere, Pierre		
; APPLICANT: Junien, Jean-Louis		
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV		
; FILE REFERENCE: 70669		
; CURRENT APPLICATION NUMBER: US/09/976,674		
; CURRENT FILING DATE: 2001-10-12		
; PRIOR APPLICATION NUMBER: US 60/240,117		
; PRIOR FILING DATE: 2000-10-12		
; NUMBER OF SEQ ID NOS: 61		
; SOFTWARE: PatentIn version 3.1		
; SEQ ID NO 36		
; LENGTH: 4180		
; TYPE: DNA		
; ORGANISM: Homo sapiens		
US-09-976-674-36		
Alignment Scores:		
Pred. No.:	2,976-134	Length: 4180
Score:	1086.50	Matches: 199
Percent Similarity:	69.25%	Conservative: 42
Best Local Similarity:	57.18%	Mismatches: 58
Query Match:	64.67%	Indels: 49
DB:	10	Gaps: 1
US-10-070-464-3 (1-310) X US-09-976-674-36 (1-4180)		
QY	2	GIUGLIThrLysAspSerProLeuGlnHisLeuTyValValSerTyValAsnPro 21
DB	1873	AAGGACCAAGAGACACCCGCTGAGACACACCTTACGTGCTAGTATGAGCGGCC 19322
QY	22	GlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHis 41
DB	1933	GCGGAGATCGTACGCTACACACGCGCGCTTCTCCATAGCTGTCCATAGACCAAGAC 1939232
QY	42	CysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyr 61
DB	1993	TTGCGACATGTTCTGTACGCCACTACAGACGCTGACACGCGCCCTCGCTCACGCTTAC 20522
QY	62	LysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIle 81
DB	2053	AAGCTGACCGGCGCCGACGACGACCCCTGCAACAAGCAGCCCGCTTGTGGCTAGCATG 21122
QY	82	LeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlnSer 101
DB	2113	ATGAGAGGACGACGACGTGCGCCCGGATTAATGTCTCCAGAGATTTTCATTCCACACG 21722
QY	102	ThrThrGlyPheThrLeuTyrGlyPheLeuTyrLysProHisAspLeuGlnProGlyLys 121
DB	2173	CGCTCGAGTGTGCGGCTCTACGCGCATGTATCAAGACCCCAAGCGCTTGACAGGAGAG 22322
QY	122	LysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
DB	2223	AAGCACCCACCGTCTCTTTGTATATGAGAGCCCCCAGGTGACAGCTGTGAATTAATCTCC 22922
QY	134	----- 134
DB	2293	TTCAAAAGCATCAAGTACTTGCGGCTCAACACACTGAGCTCGGTGCTACGCGGTGTT 23522
QY	134	----- 134
DB	2353	GTGATTGACGCGAGGGGCTCTGTTCAGCGAGGCGCTTGCGTTTGAAGGAGCCCTGAAAAAC 24122

Qy	135	-----GlyGINIIeGIIuIIeaSPaSPGInValGIuGIYLeuGINTYrLeuAlaSerArg	152
Db	2413	CAAAATGGGCGCAGGTGGAGATCGAGGACCAAGGTGAGAGGGCGTTCAGATTCGTGCGCGAAGAG	24172
Qy	153	TYrAgPheIIeaSPLeuAspArgValGIYrIleIIaGIYrTPSerTYrGIYGIYrIleu	172
Db	2473	TATGGCTTCATCGACCTTAGCCGAGTTGCCATTCATGCTGTGCTCTTAGCGGGGCTTCCTC	25332
Qy	173	SeIueMeAlaLeuMeGIaArgSeraspIIePheArgValaIIaIIealIGIYAlaPro	192
Db	2533	TCGCTCATGGGGCTATATCCACAAGGCCAGGTGTTCMAAGTGGCCATCGCGGGTCCCG	2592
Qy	193	ValThrLeuTrpIIePheTYrAspThrGIYrTYrThrGIuaGTYrMeGIYhISProasp	212
Db	2593	GTCAACCGCTGGATGGCCCTTAGCAACAGGGGTACCTGAGCCCTCATGTGAGCGTCCCTAG	26532
Qy	213	GIuaNGIuGIuNGIYTYrTYrLeuGIYSerValaIaMeGIaIIaGIuLYaPheProser	232
Db	2653	AACAAACGACGCGCTATGAGGGGTTCCGTGGCCCTGCACGTGAGAAAGCTGCCAAT	27122
Qy	233	GIuProAnaTgLeuLeuLeuIIaGIYrPheLeuAspGIuaNValIIaSPheAlaIIIS	252
Db	2713	GAGCCCAACCGCTGTCTTATCTCCACGCGCTTCCTGAGCAAAACGTGCACATTTTCCAC	2772
Qy	253	ThrSerIIeLeuLeuSerPheLeuValaArgValaGIYrProTYrAspLeuGINIIeYr	272
Db	2773	ACAAACTTCCTCGTCTCCCAACTAGTCGACAGAGGAAACCTTACCACTCCAGATCTAC	2832
Qy	273	ProGINGIuaTgHISerIIeArgValaProGIuSerGIYhIIaSTYrGIuLeuIIIShIeU	292
Db	2833	CCCAACGAGAGACACAGATTCGCTGCCGCCCAAGTCGGCGAGCACTATGAAGTCACGTTG	2892
Qy	293	LeuHISYrLeuGINIuaNleu	300
Db	2893	CTGCACCTTCTACAGAAATACCTC	2916
RESULT 14			
US-09-976-674-34	/ Sequence 34, Application US/09976674		
	/ Patent No. US20020115843A1		
	/ GENERAL INFORMATION:		
	/ APPLICANT: Qi, Steve		
	/ APPLICANT: Akinaanya, Karen		
	/ APPLICANT: Riviere, Pierre		
	/ APPLICANT: Junien, Jean-Louis		
	/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIV		
	/ FILE REFERENCE: 70669		
	/ CURRENT APPLICATION NUMBER: US/09/976,674		
	/ PRIOR FILING DATE: 2001-10-12		
	/ PRIOR APPLICATION NUMBER: US 60/240,117		
	/ NUMBER OF SEQ ID NOS: 61		
	/ SOFTWARE: PatentIn version 3.1		
	/ SEQ ID NO 34		
	/ LENGTH: 4263		
	/ TYPE: DNA		
	/ ORGANISM: Homo sapiens		
	US-09-976-674-34		
Alignment Scores:			
	Pred. No.:	3,07e-134	Length: 4263
	Score:	1086.50	Matches: 199
	Percent Similarity:	69.25%	Conservative: 42
	Best Local Similarity:	57.18%	Mismatch: 58
	Query Match:	64.67%	Indels: 49
	DB:	10	Gaps: 1
US-10-070-464-3 (1-310) x US-09-976-674-34 (1-4263)			
Qy	2	GIuGIYThrIYAspSerProIeUGIuHISleuTYrValaIaSerTYrValaIaSPPro	21
Db	1873	AAGGGCACCAAGACACGCGCTGAGACCAACTTACAGTGGTCACTATGAGGGCGGC	1932

```

QY      22  GlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHis 41
      1933  GGGGAGATGCTAGTACCGCTCAGCAGCCGCGCTCTCCCATAGCTGCTCCATGAGCCAGAAC 1992
QY      42  CysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyr 61
      1993  TTCGACATGTCCTCGACGCCATACAGACGGTAGAGACCGCCCGCGCGACGCTCTAC 2052
QY      62  LysLeuSerSerProGluAspAspProThrCysValThrLysGluPheTrpAlaThrIle 81
      2053  AAGCTAGAGCGCGCCGACAGACGACCCCTGCACAGAGCCCGCTTCGGGCTAGCAGT 2112
QY      82  LeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSer 101
      2113  ATGAGAGCAGCGAGCTGCGCCCGGATATGTTCTCTCCAGAGATCTTCCATTTCCACAG 2172
QY      102  ThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLys 121
      2173  CGCTCGAGATGTCGGCTCTAGCGCATGATCTTACAGCGCCACGCTTGACCGAGGAG 2232
QY      122  LysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
      2233  AAGCACCACCGCTCTCTTGTATATGAGAGGCCCGCCAGGTGACGCTGTGAATTACTCC 2292
QY      134  ----- 134
      2293  TTCAAAGCATCAAGTACTTGCGGCTCAACACACTGGGCTCCTCGGAGTACGCGGTGTT 2352
QY      134  ----- 134
      2353  GTGATTGACGGCAGGGGCTCTCTCTCAGCGAGGCTTCGGTTGCAAGGGCGCTGAAAAAC 2412
QY      135  -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArg 152
      2413  CAAATGGCCAGGTGGAGATCGAGGACCAAGTGAAGGGCTGCAAGTTCCGAGCGCGAAG 2472
QY      153  TyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeu 172
      2473  TATGGCTTCATCAGCTGAGCCAGGTTGCCATCGGCTGCTGCTTCAACGGGGGCTTCCTC 2532
QY      173  SerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPro 192
      2533  TCCTCATGGGGCTTATTCACAAAGCCCGAGGTGTTCAAGGTGCCATCGCGGGTGCCCG 2592
QY      193  ValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAsp 212
      2593  GTACCGCTCGATGGCTTACGACACAGGGTACACTGAGCGCTACATGAGCGTCCCTGAG 2652
QY      213  GlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPheProSer 232
      2653  AACCAACAGACACGGCTATGAGGGGCTTCGTCGCGCTTGACCGTGAAGAGCTGCCCAAT 2712
QY      233  GluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHis 252
      2713  GAGCCCAACGGCTTGCTTATCTCCACGGCTCTCTGAGCAAAAAGTGCACCTTTTCCAC 2772
QY      253  ThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyr 272
      2773  ACAAACTTCTCTCTCCCAACTGATCGAGAGGAAACCTTACCAAGCTCCGATCTAC 2832
QY      273  ProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHisLeu 292
      2833  CCCAAGAGAGACAGTATTCCTGCGCCGAGGTGCGGGCAGACATATGAAGTCACAGTGG 2892
QY      293  LeuHisTyrLeuGlnGluAsnLeu 300
      2893  CTGCACTTTCTACAGGAATACCTC 2916

```

RESULT 15  
 US-10-098-841-100  
 ; Sequence 100, Application US/10098841  
 ; Publication No. US20020197679A1

```

; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/098,841
; PRIOR FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pf_fl_genes Version 1.0
; SEQ ID NO 100
; LENGTH: 2801
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (158)..(1540)
; US-10-098-841-100

Alignment Scores:
Pred. No.: 2,71e-127 Length: 2801
Score: 1032.50 Matches: 194
Percent Similarity: 66.208 Conservative: 43
Beet Local Similarity: 54.198 Mismatches: 54
Query Match: 61.464 Indels: 67
DB: 13 Gaps: 3

US-10-070-464-3 (1-310) x US-10-098-841-100 (1-2801)
QY      1  PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSerTyrValAsn 20
      1752  TTCAGAGGACACAGACAGCGCGCTGAGCACCACTTACGTGCTCAGTATGAGGCG 1811
QY      21  ProGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
      1812  GCCGGGAGATGCTAGGCTCTCACACGCCCGGCTTCTCCATAGCTGCTCCATGAGCCAG 1871
QY      41  HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
      1872  AACTTGAATGTCGTGTCAGCCACTACACAGCGTAGAGACCGCGCGCTGCGAGCAGTC 1931
QY      61  TyrLysLeuSerSerProGluAspAspProThrCysValThrLysGluPheTrpAlaThr 80
      1932  TCAAGACTAGAGCGCCCGACAGACGACCCCTGCACAGACAGCCCGCTTCTGGGCTAGC 1991
QY      81  IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
      1992  ATGATGAGAGCAGCC-----AAGTCTTCCATTTCCAC 2024
QY      101  SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
      2025  ACGCGCTCGATGTCGGCTCTACGCGCATGATCTCAAGCCCGCCAGCTTGACAGCGAG 2084

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:24:12 ; Search time 1862.46 Seconds  
(without alignments)

4045.398 Million cell updates/sec

Title: US-10-070-464-3  
Perfect score: 1680  
Sequence: 1 FEGTKDPLEHHLVYVSYVN.....HLHLVQLNLSRIALAKVI 310

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2.1/USPRO.spool/US10070464/runat\_15102003\_113553\_24829/app\_query.fasta\_1.2652  
-DB=EST -QFMT=fastcap -SUFFIX=first -MINMATCH=0.1 -FOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10070464@cgn2.1 1 8056 @runat\_15102003\_113553\_24829 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gssr1:\*

29: gb\_gssr2:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1605.5	95.6	3143	11 AK016546	AK016546 Mus muscu
2	1605.5	95.6	5517	11 AK029788	AK029788 Mus muscu
3	1309	77.9	1042	12 BM557438	BM557438 AGENCOURT
4	1094.5	65.1	3327	11 AK050021	AK050021 Mus muscu
5	1094.5	65.1	3376	11 AK050023	AK050023 Mus muscu
6	1093.5	65.1	3457	11 AK078301	AK078301 Mus muscu
7	1052.5	62.6	789	10 BG709118	BG709118 602675382
8	1036	61.7	957	13 BG675006	BG675006 AGENCOURT
9	1021.5	60.8	1041	12 B1084090	B1084090 602869453
10	996.5	59.3	746	14 B751026	B751026 B751026
11	991	58.0	910	13 B0675260	B0675260 AGENCOURT
12	962	57.3	788	10 BG163397	BG163397 602338360
13	961.5	57.2	735	13 BU447277	BU447277 603766685
14	939.5	55.9	855	12 B1084885	B1084885 602869453
15	938	55.8	1085	9 AL582306	AL582306 AL582306
16	935.5	55.7	741	14 CB595102	CB595102 AGENCOURT
17	929	55.3	708	13 BU622228	BU622228 UI-H-FH1-
18	929	55.3	735	9 AL043338	AL043338 DKFZP4340
19	929	55.3	910	10 BG479035	BG479035 602526111
20	929	55.3	1002	10 BE733691	BE733691 601569288
21	929	55.3	1201	9 AL542617	AL542617 AL542617
22	925	55.1	776	14 CB522171	CB522171 UI-M-GH0-
23	925	55.1	800	12 BG921267	BG921267 602824107
24	925	55.1	905	14 CB209486	CB209486 AGENCOURT
25	918	54.6	746	14 BY764370	BY764370 BY764370
26	902.5	53.7	695	13 BG715866	BG715866 602675282
27	887.5	52.8	968	13 B0671635	B0671635 AGENCOURT
28	882	52.5	804	13 B0687687	B0687687 UI-CF-EC1
29	877	52.2	542	13 BX259029	BX259029 BX259029
30	864	51.4	617	10 BE789604	BE789604 601481552
31	862	51.3	674	13 BU229303	BU229303 603800110
32	855	50.9	603	12 B1455842	B1455842 603171060
33	845.5	50.3	1082	13 BU239476	BU239476 603322338
34	845	50.3	493	9 AA278625	AA278625 z878911.r
35	838.5	49.9	674	10 BE888665	BE888665 601513061
36	829	49.3	484	9 AA417787	AA417787 zV01a01.r
37	822	48.9	797	14 CA749361	CA749361 UI-M-FY0-
38	809.5	48.2	639	10 BG165205	BG165205 602344074
39	801	47.7	852	13 BU236501	BU236501 603408785
40	796	47.4	760	9 A1917735	A1917735 c11id07.x
41	792	47.1	627	10 BG752328	BG752328 602730802
42	766.5	45.6	753	9 AL040398	AL040398 DKFZP434A
43	762.5	45.4	853	12 B1223892	B1223892 602941035
44	760	45.2	897	13 BUI91638	BUI91638 AGENCOURT
45	738.5	44.0	658	10 BE388695	BE388695 601286345

## ALIGNMENTS

RESULT 1  
LOCUS AK016546 3143 bp mRNA HTc 05-DEC-2002  
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog  
ACCESSION AK016546  
VERSION AK016546.1 GI:12855334  
KEYWORDS  
SOURCE HTc; CAP trapper.  
ORGANISM Mus musculus  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636
REFERENCE AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischer, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuell, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staahl, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Botelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Wetz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohsaki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 3143)
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	6 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hatamoto, K., Hirooka, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P., Tanaka, T., Tejima, Y., Toy, T., Yamamuro, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	7 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT	Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: resgsc.riken.go.jp, URL: <a href="http://genome-gsc.riken.go.jp/">http://genome-gsc.riken.go.jp/</a> , Tel: 81-45-503-9222, Fax: 81-45-503-9216) Please visit our web site ( <a href="http://genome-gsc.riken.go.jp/">http://genome-gsc.riken.go.jp/</a> ) for further details.
FEATURES	<p>source</p> <p>1..3143</p> <p>location/Qualifiers</p> <p>organism="Mus musculus"</p> <p>mol_type="mRNA"</p> <p>strain="C57BL/6J"</p> <p>db_xref="PANTOM DB:4932434F09"</p> <p>db_xref="MGI:1907554"</p> <p>db_xref="taxon:10090"</p> <p>clone="4932434F09"</p> <p>sex="male"</p> <p>tissue_type="testis"</p> <p>clone_lib="RIKEN full-length enriched mouse cDNA library"</p> <p>dev_stage="adult"</p> <p>327..3005</p> <p>note="unnamed protein product; DIPEPTIDYL PEPTIDASE 8 homolog (Homo sapiens) (SPR1 Q9HBM5, evidence: FASTA, 95.%id, 100%length, match=2649)</p> <p>putative"</p> <p>protein_id="BAB30295.2"</p> <p>db_xref="GI:26382128"</p> <p>translation="MKIPSGRCNMAAMETQLGVEIFETACEGNGESQDRPLKEP FYERYYSQKLLADPRKYGYMAAPHDPMFKTDPSPSRDYYVAMSGEN RENTFYAEIPIKTRAAVLMVSMKPLDLDPATLDYGVSEELERKKIGTVGI AAYDVRGSGFLPQAGSGIYHIKDGSGHPQOPLRVNLTSCPNIRMDKLPAD PMVAFHNSNDIWNISLVTRERRTTYVHNELANNEEDPRAGVATFVLQSEFDYSG YWCPQARVTSGKILRLIYEENDESEVEIIVHTSPMLETRAUSFRYPKGTANPK VTFKSEIVVAAGGIIIDVDELQPEIIEGVEYIARAGMTPEGKAWGILDRS OTHQIVLISPELPIPEVDAMDROLIESVDSVPLIYEETDIIWINHIDPHF POTHEDEIEFTASECTGPRHLKYITSLIKESKYRSGGLPARSDPCKEETI TSCEWEVIGRHSNINWDEARKLYVFEETKSPLEHHLYTYSVANGPEVPLTDGYS HSCCSRCDPFIKRSYKQKPHCYSLKLSPEDDPVHKTETPATIIDSAGPLPDY TPPELTFESTFTFLVLMKPHDQGEKXKPYLVLYGQVQVQVNNRFGVKFR LNTLASIGVVVVVDNRSGCHRGKLEBAGFXKQWQIETDDEGLQYLAASDFDL DRVGIHGSYGGYSLMALMQRSDIFRVAIGAPATLMI FYDTGTERYMGHPDNEQ GYVLSGVAMQAEKPESEPNRLLLHGPIDENVHFAHTSLISFLVRAGKPYDLQYTPQ ERHSIRVPSGSHVETLHLVHLOEIGSRIALAKTY"</p>
CDS	
BASE COUNT	894 a 688 c 773 g 788 t
ORIGIN	
Alignment Scores:	1..63e-169
Pred. No.:	1605..50
Score:	84.96%
Percent Similarity:	84.12%
Best Local Similarity:	95.57%
Query Match:	11
DB:	Gaps: 1
US-10-070-464-3 (1-310) x AK016546 (1-3143)	
QY	1 PheGIuGIyThlYAsPsePrLeuEnGIuHsIstLeuTyValVaIseTyValaAn 20



Db	1926	TTTGAAGGACCAAGACTCTCTTGGAGCATCACTGTACGTCACCAAGTTATGCAAC	1985
Qy	21	ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln	40
Db	1986	CCTGGAGAAAGTGTAAGGCTGACTGACCGTGGACTACTACACTCTCGTGGCTCAGCCGG	2045
Qy	41	HisCysAspAlaPhePheIleSerLysTyrSerLysGlnLysAsnProHisCysValIleSerLeu	60
Db	2046	CATTGTGACTCTTCATATGAAGTACAGCAACCAAGAAATCCACACTGTGTGTCCCTC	2105
Qy	61	TyrllysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr	80
Db	2106	TACAAACTCTCAAGTCTCGAGAGTACCCAGTTCTATTAACCAAGAAATTTTGGCCACC	2165
Qy	81	IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu	100
Db	2166	ATTTTGGATTCAGGAGGTCTCTTCTCTGACTACACCTCCAGAAATTTTCTTTTGA	2225
Qy	101	SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly	120
Db	2226	AGTACTACTGGAATTTACACTGTATGGAATGTTGATTAAGCCCTCATGACTTACACCTGGA	2285
Qy	121	LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln	134
Db	2286	AAGAAATACCCCACTGTGTATTCTATATGTGTGTCCCCAGTGCAGCTGTGTCACAAT	2345
Qy	134	-----	134
Db	2346	CGGTTTAAAGGAGTCAAGTATTTCCGCTCGAAACACCCTGGCTCCCTGGTTATGTGTT	2405
Qy	134	-----	134
Db	2406	GTGTGTATGACAAACAGGGAGTCTGTGACCCGAGGACTTAAATTTGAAGGCGCTTTAA	2465
Qy	135	-----GlyGlnIleGluIleAspAspGlnValGlnGlyLeuGlnTyrLeuAlaSer	151
Db	2466	TATATAATGGGTCAATATGAATATCATATATCAAGTGAAGAGATCTCCAGTACTAGACTCT	2525
Qy	152	ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheTyrGlyGlyTyr	171
Db	2526	CAGTATGACTTCATTTGACTTGATGATGGAGGACATCCACGGCTGGTCTTATGTGTGCTAC	2585
Qy	172	LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheAlaGlyAlaIleAlaGlyAla	191
Db	2586	CTTCCCTCCGATGGCAATTAATACAGAGGTGGATATCTTCCGGGTGGCATATGTCTGGGGCC	2645
Qy	192	ProValThrLeuTyrPhePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro	211
Db	2646	CCAATCACTCTGTGATCTTCTATGATACAGATACACGAGCGCTATATGGTCAACCTT	2705
Qy	212	AspGlnAsnGlnGlnIleTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro	231
Db	2706	GACCAAGATGAACAGGGCTACTACTTGAATCTGTGGCCATGCAACGCGAAGATTCCCC	2765
Qy	232	SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla	251
Db	2766	TCAGAACCCMAACCGGTATCTCTCTGTGATGGGTCTTGGATGAGAAATGTTCACTTTGCA	2825
Qy	252	HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle	271
Db	2826	CACACCAATATATTCGTGATGATTTTATGTAGAGGCTGGAAGGCCATATGACTTACAGATC	2885
Qy	272	TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGlyLeuHis	291
Db	2886	TATCTTCAGAGAGGACACAGATCAGAGATTCCTGTGAGTCTGGAGAAACATTAAGACTGCAC	2945
Qy	292	LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLysValIle	310
Db	2946	CTGTCTCACTACCTTCAGAGAACCTTGGATCGGTATCGTGTGTCTGTGAAGTATATA	3002

LOCUS DEFINITION	5517 bp	RNA	linear	HTC 05-DEC-2002
AK029788	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:930560c15 product:DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens], full insert sequence.			
ACCESSION KEYWORDS	AK029788.1 GI:26081520			
ORGANISM	HTC; CAP trapper.			
REFERENCE	Mus musculus (house mouse)			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	Carninci, P. and Hayashizaki, Y.			
JOURNAL	High-efficiency full-length cDNA cloning			
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)			
PUBMED	99279253			
REFERENCE	10349636			
AUTHORS	2			
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)			
PUBMED	20499374			
REFERENCE	11042159			
AUTHORS	3			
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ozari, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
JOURNAL	RIKEN integrated sequence analysis (RISA) system -384-format			
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)			
PUBMED	11076661			
REFERENCE	4			
AUTHORS	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Glaser, C., King, B., Kochiwa, H., Kiehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Scudliff, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldaragalli, R., Barth, G., Blake, J., Boffelli, D., Boujmaa, N., Fletcher, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fritcher, C., Fujita, M., Gariboldi, M., Gasteinich, S., Hill, D., Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarella, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Saeki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Scotch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Welt, C., Whitaker, C., Wilting, L., Wyshak-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.			
JOURNAL	Functional annotation of a full-length mouse cDNA collection			
MEDLINE	Nature 409 (6821), 685-690 (2001)			
PUBMED	21085660			
REFERENCE	11217851			
AUTHORS	5			
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.			
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs			
MEDLINE	Nature 420, 563-573 (2002)			
PUBMED	6 (bases 1 to 5517)			
REFERENCE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,			

RESULT 2  
AK029788

me, T.,  
ya, T.,

TITLE  
JOURNAL

Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/  
URL: http://phantom.gsc.riken.go.jp/.

## FEATURES

## source

Location/Qualifiers

1..5517

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:4930560C15"

/db\_xref="taxon:10090"

/clone="4930560C15"

/sex="male"

/tissue\_type="testis"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

2221..4837

/note="DIPEPTIDYL PEPTIDASE 8 homolog (Homo sapiens) (SPR1) (Q9HBM5, evidence: FASTA, 95.7%id, 100%length, match=2649)

putative"

## misc\_feature

BASE COUNT  
ORIGIN

1514 a 1224 c 1230 g 1549 t

Alignment Scores:

Pred. No.: 3,91e-169 Length: 5517

Score: 1605.50 Matches: 302

Percent Similarity: 84.96% Conservative: 3

Best Local Similarity: 84.12% Mismatches: 5

Query Match: 95.57% Indels: 49

DB: 11 Gaps: 1

US-10-070-464-3 (1-310) x AK029788 (1-5517)

QY 1 PheGluGlyThrTyrAspSerProLeuGluHisLeuTyrValValSerTyrValAsn 20

DB 3758 TTGAAAGGACCAAAACCTCTCTTGGAGCATCAGCTGAGTGCACGATTATGCAAC 3817

QY 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40

DB 3818 CCGAGGAAGTGGTGGCTGAGCTGACGCGTACTACACCTCGTGGCTCAGCCG 3877

QY 41 HisCysAspPhePheLeuSerTyrSerAsnGlnHisAsnProHisCysValSerLeu 60

DB 3878 CATTGGACTCTCTCTAAGTAAGTACACCAACCAAAATCCACACTGTGTGCTCC 3937

QY 61 TyrTyrLeuSerSerProGluAspAspProThrCysTyrThrLeuGluPheTrpAlaThr 80

DB 3938 TCAAAACTCTCAAGTCTGAGATGACCCAGTTCATAAACAAAGAAATTTGGGCGACC 3997

QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100

DB 3998 ATTGGATTTCAGCAGGCTCTCTCTGACTACACCCCTCCAGAAATTTCTTTGAA 4057

QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrTyrPProHisAspLeuGlnProGly 120

DB 4058 ACTACTACTGAGATTATACATGATGAAATGTTGTATTAACCTCATGACTACCACTGGA 4117

QY 121 LysLeuTyrProThrValLeuPheLeuTyrGlyGlyProGln----- 134

DB 4118 AAGAAATACCCCACTGTATTATATATGATGATGTCGCCAGGTGACGCTGTGAACAT 4177

QY 134 ----- 134

DB 4178 CCGTTAAAGAGTCAAGTATTTCCGCTGAAACACCTGCTCCTGGGTTATGTGTT 4237

QY 134 ----- 134

DB 4238 GTGTGATATGACACAGGGATCTGTGACGAGACTTAATTGAAGCGGCTTTAA 4297

QY 135 -----GlyGlnIleGluIleAspArgGlnValGluGlyLeuGlnTyrLeuAlaSer 151

DB 4298 TATTAATGAGTCAATATGAAATGATGATGATGATGATGATGATGATGATGATGATG 4357

QY 152 ArgTyrAspPheLeuAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 171

DB 4358 CAGTATGACTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4417

QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191

DB 4418 CTCTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4477

QY 192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211

DB 4478 CCAAGTACCCCTGTGATCTCTATATATACAGATACACGAGCGGCTATATGATGATGATG 4537

QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluHisPhePro 231

DB 4538 GACCAAGATGACACAGGCTACTACTTATGATGATGATGATGATGATGATGATGATGATG 4597

QY 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251

DB 4598 TCAGAACCAACCGGTTACTCTCTGATGATGATGATGATGATGATGATGATGATGATG 4657

QY 252 HisThrSerIleLeuLeuSerPheLeuValAlaGlyValPheProTyrAspLeuGlnIle 271

DB 4658 CACACCAAGTATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4717

QY 272 TyrProGluGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291

DB 4718 TATCTCTGAGGAGGACACGATCAGATGATGATGATGATGATGATGATGATGATGATGATG 4777

QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleAlaLeuValIle 310

DB 4778 CTGCTCCACTACTTACAGAGAACTTGATGATGATGATGATGATGATGATGATGATGATG 4834

RESULT 3

BM557438 1042 bp mRNA linear EST 20-FEB-2002

LOCUS DECONT 6578992 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5466759

DEFINITION 5', mRNA sequence.

ACCESSION BM557438

VERSION BM557438.1 GI:18799430

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE NIH-MGC http://mgi.nci.nih.gov/

1 (bases 1 to 1042)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

CONTACT: Robert Strausberg, Ph.D.

Email: cga@pds-rcmail.nih.gov

Tissue Procurement: DCTD/DRP

cDNA Library Preparation: Rubin Laboratory

cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM1967 row: k column: 16  
 High quality sequence stop: 697.  
 Location/Qualifiers

## FEATURES

source

1..1042  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5466759"  
 /tissue\_type="amelanotic melanoma, cell line"  
 /lab\_host="DH10b (phage-resistant)"  
 /clone\_1ib="NH\_MGC\_42"  
 /note="Organ: Skin; Vector: pOT7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH MGC Library."

BASE COUNT 279 a 223 c 234 g 303 t 3 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 8.51e-137 Length: 1042  
 Score: 1309.00 Matches: 256  
 Percent Similarity: 78.964 Conservative: 3  
 Best Local Similarity: 78.054 Mismatches: 14  
 Query Match: 77.924 Indels: 55  
 DB: 12 Gaps: 2

US-10-070-464-3 (1-310) x BM557438 (1-1042)

OY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValSerTyrValaLn 20  
 DB 61 TTTGAAGGACCAAGACTCCCTTTAGAGCATCAGCTTACGTAGCTTACGTAAT 120  
 OY 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAlaSerGln 40  
 DB 121 CTGAGAGAGTGACAAAGCTGAGTACGCGTACTACTCATCTTCTGTCGATCAGTAC 180  
 OY 41 HisCysAspPhePheLeuSerLysTyrSerAsnGlnLysAspProHisCysValSerLeu 60  
 DB 181 CACTGTGACTCTTTATTAAGTAAGTATGTAACCAAGAAATCCACACGTGTGTCCTT 240  
 OY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 80  
 DB 241 TACAAAGCTATCAAGTCTGAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCCACC 300  
 OY 81 IleuAspSerAlaGlyProLeuProAspTyrThrProGluLulePheSerPheGln 100  
 DB 301 ATTTTGATTCAGCAGTCTCTCTCTGACTATACTCCCTCCAGAAATTTCTCTTTGA 360  
 OY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
 DB 361 AGTACTACTGAGATTACATTTATGAGATGCTCTCAACACCTCATGATCTACAGCCTGA 420  
 OY 121 LysLysTyrProThrValLeuPheLeuTyrGlyGlyProGln----- 134  
 DB 421 AAGAAATATCCATCGTGTGTTTATATATGTTGTCCTCAGAGTCAGTGTGTAATAT 480  
 OY 134 ----- 134  
 DB 481 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCCTAGCCTCTAGATTATGTGTT 540  
 OY 134 ----- 134  
 DB 541 GTAGTGAATGACACAGGGGATCTCTGTACCGAGGCGTTAAATTTGAAGCGCCTTTANA 600

OY 195 -----GlyGlnIleGluIleAspAspGlnValGlyGlyLeuGlnTyrLeuAlaSer 151  
 DB 601 TATTAATATGGCTCAATATATTAATGATCAGTGGAGAGACTCCAAATATCTAGCTTC 660  
 OY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 171  
 DB 661 CCAATATGATTTATTCATGACTTACATCGTGTGGCATCAGCGGTGGTCTTATGAGATAC 720  
 OY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191  
 DB 721 CTCTCCCGATGAGCATTAATGACAGAGTCAGATATCTCAGGATTCCTATTCGTGGGCC 780  
 OY 192 ProValThrLeuThrIlePheThrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211  
 DB 781 CAGTCATCTCTGTGATCTTCTATGATACAGATACCGGTACCGTATATAGGTCACCCCT 840  
 OY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231  
 DB 841 GACCAAGATGACAGGGCTATTACTTANGATCTGTGGCCATGACAGCAAAAGTCCCTCC 900  
 OY 232 SerGluProAsnArgLeuLeuLeuLeuHisGly-PheLeuAspGluAsnVal-HisPhe 251  
 DB 901 TCTGAACCAATATGATTTACTGCTTACATGCTTTCTCGATGAGAAATGTCCTCATTTTG 960  
 OY 251 LHisThrSer-IleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeu---- 269  
 DB 961 CACATACCCAGATTTTACTGAGATTTTNTTGTGAGGGCGCTGAAAAGCCCATATGAT 1020  
 OY 270 -----GlnIleTyrPro 273  
 DB 1021 TTTACAGCAATCTATCTCT 1038  
 RESULT 4  
 AK050021  
 LOCUS  
 DEFINITION  
 AK050021 3327 bp mRNA linear HTC 05-DEC-2002  
 Mus musculus adult male liver tumor cDNA, RIKEN full-length  
 enriched library, clone: C730003D12 product: D1PEPTIDYL PEPTIDASE 9  
 homolog [Homo sapiens], full insert sequence.  
 ACCESSION  
 AK050021.1 GI:26340743  
 VERSION  
 AK050021.1 GI:26340743  
 KEYWORDS  
 HTC, CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS  
 Carninci, P., and Hayashizaki, Y.  
 JOURNAL  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE  
 99279253  
 PUBMED  
 10349636  
 REFERENCE  
 AUTHORS  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 TITLE  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL  
 Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE  
 20499374  
 PUBMED  
 11042159  
 REFERENCE  
 AUTHORS  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, U., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,  
 Suna, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiki, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubaki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 TITLE  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multiplexed sequencer  
 JOURNAL  
 Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE  
 20530913  
 PUBMED  
 11076861  
 REFERENCE

AUTHORS	
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamane, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Boujunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohseki, S. and Hayashizaki, Y.	
TITLE	
Functional annotation of a full-length mouse cDNA collection	
JOURNAL	
Nature 409 (6821), 685-690 (2001)	
MEDLINE	
21085660	
PUBMED	
11217851	
REFERENCE	
5	
AUTHORS	
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
TITLE	
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL	
Nature 420, 563-573 (2002)	
MEDLINE	
6 (bases 1 to 3327)	
PUBMED	
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kaga, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Saito, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
TITLE	
Direct Submission	
JOURNAL	
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	
COMMENT	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.	
FEATURES	
Please visit our web site for further details.	
URL: http://genome.gsc.riken.go.jp/	
URL: http://fantom.gsc.riken.go.jp/.	
Location/Qualifiers	
1..3327	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C57BL/6J"	
/db_xref="PANTOM.DB:C730003D12"	
/db_xref="taxon:10090"	
/clone="C730003D12"	
/sex="male"	
/cell_type="tumor"	
/tissue_type="liver"	

CDS	
/clone_id="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="adult"	
176..2764	
/note="unnamed protein product: DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens) (SPTPLA147179, evidence: PASTY, 9241d, 99.61length, match=2580)	
putative"	
/codon_start=1	
/protein_id="BAC34034.1"	
/db_xref="GI:26340744"	
/translation="MCSGVSVPEOVAAGMDPTAARFCVQKXSMQTSRKSGLVSKAPRDFPVQKPDSSGPHSRRLTYLGMPTGRTSLYSELPKVYREALLSKQMLDHFQATPHGVYSRESELRERRLDVFDFSESGLFLQASNSLFLCDDGKGMVPMSPMLEIKTDCSGPRMPKICPADPAFISINNSDLVAVETEE RUTLCHQOSACVLDPNPSAGVATFVIOEPRFGCMCFPASMSEGLKTRILY EVDVSEVIVHPSPALREKRTDSRYRPTSKNPKIALKLALOTDHQKIVSCE KLRQVPSFLPFKVEYIARAGTRGKTKYMANFLDRPORLDVLPLPLFPAYSE AKRQAPARVPKRVQPFVYIEVTVNWIHDIFFFPQABQDQCFIKNBCKTGF CHLYRYVELKTKRDYDWTPEPLSPTEDEFKCPKEVALTSGMEVLTRGSKITWEO TKLVYFOGTRKDPRLSHHLVYVSYEAGEIVRLTLTGFSHSCMSQSFDMFVSHSVS TPCVHVYKLGGDDPLHKOPRFWASMEANCPDPVPEIPIFHTRADVOLYGM IYKPHTLQPGRKHTVLFVYGGPOVOLVNNSPFGIKYLRNLTLASIGVAVVLDGRSC OGCHREGALKNQMGQVEIEDVEGLQYAEKYGPIDSLRAVTHGMSYCGPLSIMGLI HKQVFKVAIAAPVTVMAVDGTERFMDVPENNQGYEAGVALHKEKLPNEPR LILHGFIDENVHFHTNPLVSQLIRAGKPYOLQIYPNRHSIRCKESEGEHVEVTLH FLOEHL"	
BASE COUNT	
702 a 1015 c 919 g 691 t	
ALIGNMENT SCORES	
Pred. No.: 8,33e-112 Length: 3327	
Score: 1094.50 Matches: 200	
Percent Similarity: 69.34% Conservative: 42	
Best Local Similarity: 57.31% Mismatches: 58	
Query Match: 65.15% Indels: 49	
DB: 11 Gaps: 1	
US-10-070-464-3 (1-310) x AK050021 (1-3327)	
1 PheGluGLYThrLysAspSerProLeuGluNIshLseuTYrValIseTYrValaen 20	
1715 TTTCAGAGTACAAAGACACACCGCTGAGAACATCCTTATGTGTACGTACAGTCA 1774	
21 ProGluGLuValThrArgLeuThrAspArgGLYrSerHisSerCyAsylIseSerGin 40	
1775 GCAGGGAGATGCTGGCGCTCACACGCTCGGCTTCTCCACAGCTGCATGACCG 1834	
41 HisCyAspPhePheIleSerIYrSerArgGlnLYAsnProHisCyValIseLeu 60	
1835 AGCTTCGACATGTCGTGAGTACCTACACAGCTGACACCGCCCTGTGTACATG 1894	
61 TYrLYsLeuSerSerProGluAspAspProThrCYsLYeThrLYeGluPheTpaIaThr 80	
1896 TACACACTAGAGGCGCCCGCATATATACCATCGACACACACCGCTTCTGGGCAC 1954	
81 IleLeuAspSerAlaGLYrProLeuProAspTYrThrProGluNIlePheSerPheGlu 100	
1955 ATGATGAGAGGACCAATGACCCCGACGACTATGAGCCCTTGAGATCTTCACACTTCCAC 2014	
101 SerThrThrGlyPheThrLeuTYrGLYrGLYrGLYrGLYrGLYrGLYrGLYrGLYr 120	
2015 ACCCGTCAGACGTGACGTCTACGGCATGACCTACMACCCACACACCTCGAACCTGG 2074	
121 LysLYsTYrProThrValLeuPheIleTYrGLYrGLYrProGluNIlePheSerPheGlu 134	
2075 AGGAAGACACCCACACTGTGTCTTTGTCTATGGGGGCCACACAGTGCAGTGTGGAACAC 2134	
134 ----- 134	
2135 TCCTTTAAGGCGATCAATACCTGCGGCTAAATACACTGCGATCTTGCGGCTATGCTG 2194	
134 ----- 134	

Db	2135	GTGGGATCATGATGTCGGGGGCTCTCGTCAAGCGGGGCGTCACTTTCGAGAGGGGCGCTGA	2255
Qy	135	-----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer	151
Db	2255	AATCAAAATGGGCGAGGTGGAGATTGAGGACCGAGGTGAAGGCTTGACATACCTGGCTTAG	2314
Qy	152	ArgTyrAspPheIleAspLeuAspArgValGlyIleHisgLyTyrSerTyrGlyGlyTyr	171
Db	2315	AAGTATGGCTTCATTGACTTGAAGCCGAGTCCGACATCCATGGCTGGTCTTACCGCGGCTTC	2374
Qy	172	LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla	191
Db	2375	CTCTCACTCATCGGGCTCATTCACAAGGCAACAAGTCTCAAGGTAGCCATTGCGGGCGCT	2433
Qy	192	ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro	211
Db	2435	CTGTGCACGTGTGATGAGCCCTATGACACAGGGTACACAGGAAAGATACATGATGATGCC	2494
Qy	212	AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluIlePhePro	231
Db	2495	CAAAATTAACCGACAAAGGCTATGAGGACAGGCTTGTAGCCCTGCATGTGAGAAAGCTGCC	2554
Qy	232	SerGluProAsnArgLeuLeuLeuLeuHisgLyPheLeuAspGluAsnValHisPheAla	251
Db	2555	AATGAGCCCTAACCGCTGCTATTCTCCACAGGCTTCTTGAGACGAAAGCTTCACTTCTTC	2614
Qy	252	HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIleProTyrAspLeuGlnIle	271
Db	2615	CACACAAATTTCTCGGTGTCCCGACGTGATCCGACGACGAAAGCCATACCACTTCAGATTC	2674
Qy	272	TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHis	291
Db	2675	TACCCAAACGAGACATACATCCCTGCGCCGCGCGAGCTCGGAGACATTACGAGGTGACG	2734
Qy	292	LeuLeuHisTyrLeuGlnGluAsnIleu	300
Db	2735	CTGCTGCACCTTCTCGCAGGAACACCTG	2761
RESULT 5			
AK050023			
LOCUS			
DEFINITION	AK050023	3376 bp	mRNA
			linear
			HTC 05-DEC-2002
			Mus musculus adult male liver tumor cDNA, RIKEN full-length
			enriched library, clone:G73003H19 product:DIPEPTIDYL PEPTIDASE 9
			homolog (Homo sapiens), full insert sequence.
ACCESSION	AK050023		
VERSION	AK050023.1	GI:26093756	
KEYWORDS			
SOURCE			
ORGANISM			Mus musculus
			Mus musculus (house mouse)
REFERENCE			
AUTHORS			Bukacina, P., Wecazoa, Chordata; Carninci, P., Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS			1
TITLE			Carninci, P. and Hayashizaki, Y.
JOURNAL			High-efficiency full-length cDNA cloning
MEDLINE			Meth. Enzymol. 303, 19-44 (1999)
PUBMED			99279253
REFERENCE			10349636
AUTHORS			2
AUTHORS			Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
			Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
			Normalization and subtraction of cap-trapper-selected cDNAs to
			prepare full-length cDNA libraries for rapid discovery of new genes
			Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL			
MEDLINE			
PUBMED			10499374
REFERENCE			11042159
AUTHORS			3
AUTHORS			Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
			Konno, H., Akiyama, U., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M.,
			Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
			Yamanoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
			Fujiwaka, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Matsubiki, M.,
			Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,
			Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

TITLE	RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	
AUTHORS	4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arikawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamaneke,I., Saito,T., Okazaki,Y., Gotohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Flisbachman,W., Gaasterland,T., Gissi,C., King,B., Kochiya,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaudo,I., Pesole,G., Quackenbush,J., Schriml,L.M., Stabill,F., Suzuki,R., Tomita,M., Wagner,L., Mashio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,P., Baresch,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Butl,C., Fletcher,C., Fujita,M., Gariboldi,M., Gietzung,S., Hill,D., Hotmann,M., Hume,D.A., Kamita,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzaelli,J., Mommaerts,P., Nordone,P., Riing,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-Oka,K., Wang,K.H., Weitz,C., Whitaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohetsuki,S. and Hayashizaki,Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	02108560
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 3376)
PUBMED	Adachi,J., Aizawa,K., Akimura,T., Arikawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hixmoco,K., Hirokawa,T., Hirozane,T., Hori,F., Imocani,K., Iehi,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kochi,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koude,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,Y., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Saegaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Mutsaers,M., and Hayashizaki,Y.
REFERENCE	Direct Submission
AUTHORS	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehito-cho, Tsutsumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL:http://genome-gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
TITLE	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
JOURNAL	Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.
MEDLINE	Please visit our web site for further details.
PUBMED	URL:htp://genome-gsc.riken.go.jp/ URL:http://fantom-gsc.riken.go.jp/ Location/Qualifiers
REFERENCE	1. 3376 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J"
AUTHORS	FEATURES SOURCE

```

/db_xref="FANTOM_DB:C730003H19"
/db_xref="taxon:10090"
/clone="C730003H19"
/sex="male"
/cell_type="tumor"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
235_2824
/notes="DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens)
(SPTRI/AL47179, evidence: FASTY, 92%id, 99.6%length,
match=2580)
putative"
misc_feature
BASE COUNT      698 a      1046 c      932 g      700 t
ORIGIN
Alignment Scores:
Pred. No.:      8,53e-112      Length:      3376
Score:          1094.50      Matches:      200
Percent Similarity: 69.34%      Conservative: 42
Best Local Similarity: 57.31%      Mismatches: 58
Query Match:     65.15%      Indels:      49
DB:              11      Gaps:        1
US-10-070-464-3 (1-310) x AK050023 (1-3376)
QY      1 PheGluGlyThrIleuAspSerProLeuGluH1eHisLeuTyValSerTyValAsn 20
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1775 TTTCAGAGTACAAAGACACACCGCTGGAACATCCTTATGTGCTACCTACGAGTCA 1834
QY      21 ProGluGluValThrIleuThrIleuThrIleuThrIleuThrIleuThrIleu 40
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1835 GCAGGCGAGATCGGGCGGCTCACACGCTCGGCTTCCACAGCTGCTCCATGAGCAG 1894
QY      41 HisCysAspPhePheIleSerIleSerIleSerIleSerIleSerIleSerIle 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1895 AGCTTCGACATGTTCTGAGTCACTACAGAGTGTGAGCAGCCCTTGTACATGTG 1954
QY      61 TyrIleuSerSerProGluAspAspProThrCysIleuThrIleuGluPheIleu 80
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1955 TACAGAGTACGCGCCCGCATGATGACCTGACAGACAGACACCGCTTCTGGCGCAGC 2014
QY      81 IleuAspSerIleuIleuProLeuProAspTyrThrProGluIlePheSerPheGlu 100
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2015 ATGATGAGGACGACCAATTCGCCCCGACATGCTGCCCCCTGAGATCTTCCACTTCAC 2074
QY      101 SerThrThrGlyIleuThrIleuThrIleuThrIleuThrIleuThrIleuThr 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2075 ACCCGTGCAGACGTCGACGCTCTACGGCATATCTACAGACACACCTGCAACCTGGG 2134
QY      121 LysIleuTyProThrValIleuPheIleTyGlyGlyProGlu 134
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2135 AGGAGAGACCCCACTGCTCTTGTCTATGGGGGCCCAAGGCGAGTGTGAACAC 2194
QY      134 ----- 134
DB      2195 TCCTTAAGGGCATCAATACTGCGGCTAAATACACTGGCATCTTGGCTATGCTGTG 2254
QY      134 ----- 134
DB      2255 GTGGTATGATGATGCGGGCTCTCTGTCAGCGGGGCTGCACCTTCAGGGGGCCCTGAAA 2314
QY      135 -----GlyGluIleGluIleAspAspGlnValGluIleuGlnIleuGlnIleu 151
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2315 AATCAATATGGCGGCGAGATGAGACCAAGTGAAGCTTGCGATCGTGGCTGAG 2374
QY      152 ArgTyrAspPheIleuAspAspValGlyIleHisGlyTTPSerTyrGlyGlyTyr 171
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2375 AAGTATGCTTCACTGATGACCGGATGCCATTCATGCTGCTTACGGCGGCTTC 2434
QY      172 LeuSerIleuMetIleuMetGlnArgSerAspIlePheArgValAlaIleIleGlyAla 191
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2435 CTCTCACTCATGGGGCTCATCCACCAAGCCCAAGTGTTCAGGTAGGCACTTGGCGGCT 2494

```

```

QY      192 ProValThrIleuThrIleuPheTyAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2495 CCTGTACTGTGTGATGGCGCTATGACACAGGTTACAGGAAGATACATGATGTCCTCC 2554
QY      212 AspGlnAsnGluGlnGlyTyrTyrIleuGlySerValAlaMetGlnIleuGlyPhePro 231
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2555 GAAATATACAGCAGCAAGCATATGAGGAGGCTGTGATCCCTGATGTGAGGAAGCTGCC 2614
QY      232 SerGluProAspAspIleuLeuLeuIleGlyPheLeuAspGluAsnValHisPheAla 251
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2615 AATGAGCTTACCGCTGCTTATCTCCACGCTTCTGGACGAGAACGTTCTCTTC 2674
QY      252 HisThrSerIleuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2675 CACACAATTTCTGCTGTCCTCCAGCTGATCCGAGCAAGGACATACCGCTTCAGATC 2734
QY      272 TyrProGluGluArgHisSerIleArgValProGluSerGlyGluHisTyrGlyLeuHis 291
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2735 TACCAACAGCAGAGACATGATGATCCCTGCGGAGTCCGAGAGCATTTACGAGTGACG 2794
QY      292 LeuLeuHisTyrIleuGlnGluAsnLeu 300
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      2795 CTGCTGCACTTCTGCAAGGACACCTG 2821
RESULT 6
AK078301
LOCUS
DEFINITION
MUS musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430584g1 product:DIPEPTIDYL PEPTIDASE 9
homolog [Homo sapiens], full insert sequence.
ACCESSION
AK078301
VERSION
AK078301.1 GI:26347124
KEYWORDS
HTC; CAP trapper.
MUS musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279259
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugihara, Y., Shibata, K.,
Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadoya, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T.,
Fleischmann, M., Gasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,

```

Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bulc, C., C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wyszynski, B., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S., and Hayashizaki, Y.

TITLE  
JOURNAL MEDLINE  
PUBLISHED  
REFERENCE  
AUTHORS

Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3457)  
Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, D., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasakawa, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE  
JOURNAL

Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

FEATURES  
source

CDS

1..3457  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:6430584G11"  
/db\_xref="taxon:10090"  
/clone="6430584G11"  
/sex="male"  
/tissue\_type="Olfactory brain"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
182..2770  
/note="unnamed protein product, DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens) (SPTPLA47179, evidence: FASTY, 92%ID, 99.6%length, match=2580)  
putative"  
/codon\_start=1  
/protein\_id="BAC37211.1"  
/db\_xref="GI:26347125"  
/translation="MCSGVSPVEQVAAQMDMTAARFCVQKHSMDGLSIIHSGSKSS GLIVSKAPHDFOFVQKPDSEGPSHRLYYLQMPYGSRENSLLSEYIPKKVRKALLLL

BASE COUNT  
ORIGIN  
766 a 1033 c 926 g 732 t

Alignment Scores:  
Pred. No.: 1,15e-111 Length: 3457  
Score: 1093.50 Matches: 200  
Percent Similarity: 69.34% Conservative: 42  
Best Local Similarity: 57.31% Mismatches: 58  
Query Match: 65.09% Indels: 49  
DB: 11 Gaps: 1

US-10-070-464-3 (1-310) x AK078301 (1-3457)

QY 1 PHEGLGLYTHLYAHPSEPRLEUGLHSHLSLEUTYVALVALSERTYVALA 20  
DB 1721 TTTCAGGATGACAAAGACACACCGCTGGAACATCACTTATGCTGACGATCA 1780  
QY 21 PROGLYGLVALTHARGLEUTHRAAPARGLYTYRSEHISSERCYAGSILESERGIN 40  
DB 1781 GCAGGAGATGCTGGCTGCTCACCGCTGCTTCCACAGCTGCTCCATGACCCAG 1840  
QY 41 HISCYAASPHEPHEIIESELYTYRSEARANGINLYAASPHEHISCYASVALSERLEU 60  
DB 1841 ACCCTGACATGTTGCTGATGCTACACAGCATGAGACACCGCTGATGATGAG 1900  
QY 61 TYRGLYGLSERSEPRGILUASPAPROTHCYSLYETHLYAGLUBHETRALATHR 80  
DB 1901 TCAAGAGTGAAGGGGCCCCCATGATATGACCCACGACCAACCAACGCTTCTGGGCGAC 1960  
QY 81 ILEUAAPSERVALGLYPROLEUPROAAPTYYTHPRPROGLIUILEPHESERPHGLU 100  
DB 1961 ATGATGAGGACGCAATGTCCTCCAGACGTAATGCCCCCTGAGATCTTCCATTCAC 2020  
QY 101 SERTHTRHGLYPHEHTRLEUTYRGLMETLEUTYRGLPHEHISAPLEUGLIPROGLY 120  
DB 2021 ACCCGTGCAGAGTCGACGCTCTACGCGATGATCTACAGCCACACCCCTGCACTGG 2080  
QY 121 LYSLYETYPRTHTVALLEUPHEILETYRGLYGLYPROGLN----- 134  
DB 2081 AGGAAGACCCACGACGCTGCTTGTCTATGAGGGGCCACAGGTGACGATGTGAACAAC 2140  
QY 134 ----- 134  
DB 2141 TCTTTAAGGATCAATATCTGCGGCTAATAATACATGGCATTCCTGGGCTATGCTG 2200  
QY 134 ----- 134  
DB 2201 GTGGATGATGATGTCGGGCTCTGTACGCGGGGCTGCACTTGAGGGGCGCTGA 2260  
QY 135 -----GLYGLIIEGLIUIEAPAPGINVALGLUGLYLEUGLINTYRLEUALAS 151  
DB 2261 AATCAATGCGGACAGGTGAGTTGAGACCGAGGTGAGGCTTGACATGCTGAGTGA 2320  
QY 152 ARGTYRAAPHEIIEAPLEUASPARGVALGLYIIEHISGLYTRPSERTYRGLYGLY 171  
DB 2321 AAGTATGCTTATGATCTGAGCCGAGTCCGACATCCATGCTGAGTCCACGCGGCTTC 2380  
QY 172 LEUSERLEUALLEUMERGLNARGSERAPILIEHARGVALAIIIEALAGLYA 191



Db 2381 CTCCTACGATGGGCTCATCCAGCAAGCTTCAGGTACCATTCGGGGCGCT 2440  
 QY 192 ProvalThleutrpilPhePheTyAspThrGlyTyThrGluArgTyMetGlyHisPro 211  
 Db 2441 CCTGTACTGTGTGGATGGCTTATGACACAGGGTATACCGAAGATATGATGTCCCC 2500  
 QY 212 AspGlnAsnGluGlnGlyTyTyTyLeuGlySerValAlaMetGlnAlaGluTySphPro 231  
 Db 2501 GAAATATACACAGCAAGGCTATGAGCGAGGCTGTGAGCCCTGCATGTGAGGAAGCTGCC 2560  
 QY 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisSphAla 251  
 Db 2561 AATGAGCCCTAACCGCTGCTTATCTCCAGCGCTTCCTGAGCAGAAACGTTCTTCTTC 2620  
 QY 252 HistHSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAspLeuGlnIle 271  
 Db 2621 CACACAAATTTCTGGTGTCCAGCTGATCCGAGCAGGAAGCATATCCAGCTTCGATC 2680  
 QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyTyGluLeuHis 291  
 Db 2681 TACCCAAAGAGAGACATAGCATCCGCTGCCGAGAGTCCGAGAGCATTAAGAGTGACG 2740  
 QY 292 LeuLeuHisTyTyLeuGlnGluAsnLeu 300  
 Db 2741 CTGCTGCACTTCTTGCAAGAACACTTG 2767

RESULT 7 789 bp mRNA linear EST 07-MAY-2001  
 BG709118 602675382F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4797668 5',  
 LOCUS mRNA sequence.  
 DEFINITION BG709118.1 GI:13987136  
 ACCESSION EST.  
 VERSION BG709118.1  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 789)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaos-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkevics, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LHAM10684 row: d column: 21  
 High quality sequence stop: 784.  
 Location/Qualifiers  
 1..789  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4797668"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_96"  
 /note="Organ: brain; Vector: pBluescriptR (modified  
 pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gscgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',  
 size-selected for average insert size 2.3 kb and  
 normalized to Rot 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NHGRI/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."

FEATURES  
 source  
 207 a 163 c 187 g 232 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 4..9e-108 Length: 789  
 Score: 1052.50 Matches: 208  
 Percent Similarity: 80.00% Conservative: 0  
 Best Local Similarity: 80.00% Mismatches: 2  
 Query Match: 62.65% Indels: 51  
 DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x BG709118 (1-789)  
 QY 77 PheTrrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyThrProGluIle 96  
 Db 10 TTTTGGGCCACCATTTTGGATTGACGAGGCTCTTCTGACTATATCTCTCCAGAAATT 69  
 QY 97 PheSerPheGluSerThrThrGlyPheThrLeuTyGlyMetLeuTyTySphProHisAsp 116  
 Db 70 TTTCTTTTGAAGAGTACTAGATTTTACATTTGTATGAGAGTGTCTTACAGAGCTCATGAT 129  
 QY 117 LeuGlnProGlyLysLysTyTyProThrValLeuPheIleTyTyGlyTyProGln----- 134  
 Db 130 CTACAGCCTGGAAGAAATATCTTACTGTGCTGTTCATATATGTGTCTCTAGAGTGACG 189  
 QY 134 ----- 134  
 Db 190 TTGGTAAATATCGTTTAAAGAGTCAAGATATTTCCGTTGAATACCTAGCCTCTCTA 249  
 QY 134 ----- 134  
 Db 250 GGTATATGTGTTGTAGTATAGCAACAGGGATCTGTCCAGGAGCTTAAATTGAA 309  
 QY 136 -----+-----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGln 147  
 Db 310 GCGGCTTAAATATATAAATATGAGTCAATATGAAATTGACATCAGGTGGAAGACTCCA 369  
 QY 148 TyrLeuAlaSerArgTyTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyPser 167  
 Db 370 TATCTAGCTTCTCGATATATATTCATTTGATGATCGTGTGGGCAATCCAGGCTGTCTC 429  
 QY 168 TyrGlyGlyTyTyLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAla 187  
 Db 430 TATGAGAGATACCTCTCTCCGATGCAATATGACAGAGGTCAAGATATCTTCAGGGTTGCT 489  
 QY 188 IleAlaGlyAlaProValThrLeuTrpIlePheTyAspThrGlyTyThrGluArgTyTy 207  
 Db 490 ATTGCGGGGGCCCCAGTCACTCTGTGATCTTATGATATACAGATATACAGAAACGTTAT 549  
 QY 208 MetGlyHisProAspGlnAsnGluGlnGlyTyTyTyLeuGlySerValAlaMetGlnAla 227  
 Db 550 ATGGGTACCCCTGACCTGATGATGAACAGGGCTATTACTTAGAGATCTGTGGCATGCAAGCA 609  
 QY 228 GluTySphProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsn 247  
 Db 610 GAAAGATTCCCTCTTACCAAAATCGTTTACTGCTCTTACATGCTTCCGATGAGAAAT 669  
 QY 248 ValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTy 267  
 Db 670 GTCCATTTTGCACATACAGATATATTACTGAG- TTTTATGTAGAGGCTGGAAGAACATAT 728  
 QY 267 rAspLeuGlnIleTyTyProGlnGluArgHisSerIleArgValProGluSerGlyGlu 286  
 Db 729 TGATTTACAGATCTATCTCAGAGAGACACAGCATTAAGAGTCCCTGAATCGGAGAA 786

RESULT 8 957 bp mRNA linear EST 15-JUL-2002  
 BG675006 5', mRNA sequence.  
 DEFINITION AGENCOURT\_9303592 NIH\_MGC\_102 Homo sapiens cDNA clone IMAGE:6274622  
 LOCUS BG675006  
 ACCESSION BG675006  
 VERSION BG675006.1 GI:21785840  
 KEYWORDS  
 SOURCE Homo sapiens (human)



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 957)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 AUTHORS Unpublished  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov> row: h column: 15  
 Plate: L10C2456  
 High quality sequence start: 4  
 High quality sequence stop: 624.  
 Location/Qualifiers

FEATURES  
 source 1..957

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6274622"  
 /tissue\_type="epidermoid carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH-MGC\_102"  
 /note="Organ: salivary gland; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Library constructed  
 by Ling Hong in the laboratory of Gerald M. Rubin  
 (University of California, Berkeley) using ZAP-cDNA  
 synthesis kit (Stratagene) and SuperScript II RT (Life  
 Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 257 a 206 c 228 g 266 t

ORIGIN

Alignment Scores:  
 Pred. No.: 4,81e-106 length: 957  
 Score: 1036.00 Matches: 200  
 Percent Similarity: 76.81% Conservative: 2  
 Best Local Similarity: 76.05% Mismatches: 9  
 Query Match: 61.67% Indels: 52  
 DB: 13 Gaps: 1

US-10-070-464-3 (1-310) x B0675006 (1-957)

Oy 1 PheGluGlyThrLysAspSerProLeuGluHisIleuTyValValSerTyValAsn 20  
 Db 162 TTGAGAGCACCAGCAAGCTCCCTTTAGAGCATCCTGTACGTAGTACGTAAT 221  
 Oy 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGaillesergin 40  
 Db 222 CCGAGAGAGTGACAAAGCTGACGCTGCTACTCATTCTTGCTGATCAGTACG 281  
 Oy 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysAlaSerLeu 60  
 Db 282 CACTGTGACTCTTTATAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGT 341  
 Oy 61 TyrLysLeuSerSerProGluLysAspProThrCysLysThrLysGluPheThrAlaThr 80  
 Db 342 TACAGAGTATCAAGCTCGAAGATGACCACTTCGAAACAAAGAAATTTTGGCCACC 401  
 Oy 81 IleuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100  
 Db 402 ATTGTGATTGAGCAGGCTCTCTCTCTGACTACTCTCTCCGAAATTTTCTTTGAA 461  
 Oy 101 SerThrTrnGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120  
 Db 462 AGTACTACTGATTTACATTTGATGAGATGCTCTCAAGCCCTCATGATCTACAGCTGGA 521

Oy 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134  
 Db 522 AAGAAATATCCACTGTGTCTTATATATGTTGCTCCTCAGTGACGTGGATATAT 581  
 Oy 134 ----- 134  
 Db 582 CGGTAAAGAGTCAGATATTCGGCTTGATATACCCCTCTCTAGTTATGCT 641  
 Oy 134 ----- 134  
 Db 642 TGTACTGATAGCAAAACAGGGGATCTCTGACCGAGGCTTAATTGAGCGCCCTTT 701  
 Oy 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeu 150  
 Db 702 TAATATAAATGGGTCAATAGAAATGACATACAGGGGAGGAAAGACTCAATATTCAC 761  
 Oy 150 AserArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPseTyrGly 170  
 Db 762 TTCTGATATGATTTTCATTGACTTAGATCGTGTGGCATCCAGCGCTGCTCTATGAGG 821  
 Oy 170 YTrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAla 190  
 Db 822 ATCCCTCTCCCGAAGGATTTATGCAAGGTCAGATATCTTCCAGGTTGCTATTGCTGG 881  
 Oy 190 YAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGly 210  
 Db 882 GGGCCCATGACTCCGCTTGATCTTCATGATACAGATACCCGGAACGGTATTATGGGG 941  
 Oy 210 IsPro 211  
 Db 942 CCCCC 946

RESULT 9  
 LOCUS B1084090 1041 bp mRNA linear EST 20-JUN-2001  
 DEFINITION 602869453r1 NIH-MGC\_102 Homo sapiens cDNA clone IMAGE:5013996 5',  
 mRNA sequence.  
 ACCESSION B1084090  
 VERSION B1084090.1 GI:14502420  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1041)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: InCyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L10C1820 row: j column: 13  
 High quality sequence stop: 847.  
 Location/Qualifiers

FEATURES  
 source 1..1041

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5013996"  
 /tissue\_type="epidermoid carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH-MGC\_102"  
 /note="Organ: salivary gland; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Library constructed  
 by Ling Hong in the laboratory of Gerald M. Rubin

(University of California, Berkeley) using ZAP-cDNA  
synthesis kit (Stratagene) and Superscript II RT (Life  
Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 280 a 227 c 242 g 292 t

## ALIGNMENT SCORES:

Alignment Scores:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Pred. No.:	2,37e-104	1041	216	6	4
Score:	1021.50	Conservative:	6		
Percent Similarity:	73.03%	Mismatches:	16		
Best Local Similarity:	71.05%	Indels:	6		
Query Match:	60.80%	Gaps:	4		

US-10-070-464-3 (1-310) x B1084090 (1-1041)

```

QY      1 PHEGLUGLYThrIysAspSerProLeuGluHisIleuTyrtValValSerTyrtValaAn 20
        |||
DB      153 TTTGAAGGACCAAAACCTCCCTTTAGAGACATCAGTCTAGTACGTTACGTTAAAT 212
        |||
QY      21 ProGLGLValThrtArgLeuThrAspArgGLYTYrSerHisSerCysCysIleSerGln 40
        |||
DB      213 CCGAGAGAGGTGACAAAGGCTGAGTACCGGCTACTACATTCCTGCTGCATCATGAG 272
        |||
QY      41 HisCysAspPhePheIleSerIlyrSerAsnGlnLysAsnProHisCysValSerLeu 60
        |||
DB      273 CACTGAGACTCTCTTTAAGTAGTAGTAGTAAACAGACATCACACTGCTGCTCCCT 332
        |||
QY      61 TyrtIysLeuSerSerProGluAspAspProthCysIlyrThrtIysGluPheTpaIatTr 80
        |||
DB      333 TACAACTATCAAGTCTCTGAGAGTGAACCCAACTGCAAAAAGAAATTTTGAGCCACC 392
        |||
QY      81 IleLeuAspSerIaGLyProLeuProAspTyrtThrProGluIlePheSerPheGlu 100
        |||
DB      393 ATTTTGATTCAGCAGAGTCTCTCTGACTATACCTCCCAAAATTTTCTTTTGA 452
        |||
QY      101 SerThrThrGlyPheThrLeuTyrtGlyMetLeuTyrtIysProHisAspLeuGlnProGly 120
        |||
DB      453 AGTACTACTGAGATTTCATTTGATGGAGATGCTCTACAGAGCTCATGATACAGCTCGA 512
        |||
QY      121 LysIlyrTyrtProthrtIleuPheIleTyrtGlyGlyProGln----- 134
        |||
DB      513 AAGAAATATCTACTGCTGCTGCTCATATATGATGCTGCTCAGGTGAGTGTGAATTAAT 572
        |||
QY      134 ----- 134
        |||
DB      573 CGGTTTACAGAGAGTCAAGTATTTCGGCTGAATACCTAGCCTCTAGATTATGTGG 632
        |||
QY      134 ----- 134
        |||
DB      633 TTGTAGTATAGACAAAGGGAGTCTGTCCAGAGGGCTTAATTGAAGGGCCCTTTA 692
        |||
QY      135 -----GlyGlnIleGluIleAspAspGlnValGlyGlyLeuGln-TyrtIleuIle 151
        |||
DB      693 ATATACAAAGGTCATATGAATATGACATCAGTGTGAAGAGATCCAAATTTATTTACCTTC 752
        |||
QY      151 rArgTyrtAspPheIle-AspLeuAspArgValGlyIleHisGlyTyrtPseTyrtGly-Gly 170
        |||
DB      753 TCGATATGATTCATTCATTTGACTTATGATCGTGTGGCATCCAGGGTGTCTATGAGAGGA 812
        |||
QY      171 TyrtLeuSerLeu-MetAlaLeuMetGlnArgSerAsp-IlePheArgValAlaIleAla- 189
        |||
DB      813 TACCTCTCCCTGAATGAGCATTAATGAGAGGTAGATGAGTACAGGGTGCATATGAGCG 872
        |||
QY      190 GlyAlaProValThrLeuTrpIlePhe-TyrtAspTrnGlyTyrtThrGluArgTyrtMetG 209
        |||
DB      873 GGGGGCCCACTGCTGTGATCTCTTATGATACAGATACAGGAACTTATATGAG 932
        |||
QY      209 yHisProAspGlnAsnGlnGlnGlyTyrtTyrtIleuGlySerValAlaMetGlnIaGly 229
        |||
DB      933 TACCTGAGACAGA---TGAACGGGTCAATCTTCAGGATCTGTG-----CTGTGACAGAA 983
        |||
QY      229 sPhePro-----SerGluProAsnArgLeuLeuLeuLeuHis 241
        |||

```

DB 984 ATTCCTTGTGACAAATGCTTAAATGCTCAATGTTCCGGGCAAAATTCCTTGACATACGAC 1041

RESULT 10  
LOCUS  
DEFINITION

BY751026 BY751026 746 bp mRNA linear EST 17-DEC-2002  
CDNA clone F830018D07 5', mRNA sequence.

ACCESSION  
BY751026  
BY751026.1 GI:27181193

KEYWORDS  
EST.  
Mus musculus (house mouse)

SOURCE  
MUS MUSCULUS

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 746)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,

Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,

Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,

Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,

Geisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani,

L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,

A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,

Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,

Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,

King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,

P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,

H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Petrea, G.,

Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,

Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J. J., Ring,

B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C. A., Setou,

M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,

R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,

Watanabe, Y., Wells, C., Wilming, L. G., Wynchaw-Boris, A., Yanagisawa,

M., Yang, T., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A.,

Carrincci, P., Hayatsu, N., Hirozane-Kitahara, T., Konno, H., Nakamura,

M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,

Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,

Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,

K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,

E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Science Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-reseqc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carrincci, P., Fukuda,

S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,

Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,

H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,

Numazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N., Sano, H.,

Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Shigemitsu, Y., Takeda, Y.,

Waki, K., Watanabe, M., Muramatsu, M. and Hayashizaki, Y. Direct

Submission

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

#### FEATURES

source 1..746  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="NOD"  
 /db\_xref="taxon:10090"  
 /clone="F83001BD07"  
 /tissue\_type="activated spleen"  
 /clone\_lib="RIKEN full-length enriched, activated spleen"  
 BASE COUNT 194 a 178 c 176 g 197 t 1 others  
 ORIGIN

#### Alignment Scores:

Pred. No.: 9 386-102 Length: 746  
 Score: 996.50 Matches: 189  
 Percent Similarity: 77.73% Conservative: 3  
 Best Local Similarity: 76.52% Mismatches: 5  
 Query Match: 59.32% Indels: 50  
 DB: 14 Gaps: 1

US-10-070-464-3 (1-310) x BY751026 (1-746)

Qy 17 SerTyrValAsnProGlyValValThrArgLeuThrAspArgGlyTyrSerHisSerCys 36  
 Db 1 AGTTATGCAAAACCTCGAGAAAGTGGTGAAGGTGAGTGAACCGTGGCTACTACACCTCTCC 60  
 Qy 37 CysHisSerGlnHisCysAspPhePheHisSerIleSerIleSerAsnGlnIleAsnProHis 56  
 Db 61 TGCCCTACGCGGCAATTTGATCTCTTCATTAAGTAAAGTACAGCAACGAGAAATCCACAC 120  
 Qy 57 CysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIleThrIleGlu 76  
 Db 121 TGTGTGTCTCTTACAAACTCTCAAGTCTCTGAGATGACCCGCTTCATTAACAAAGGAA 180  
 Qy 77 PheTTPAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIle 96  
 Db 181 TTTTGGGCAACATTTGGATTACAGAGTCTCTTCTGACTACACCCCTCCAGAAAT 240  
 Qy 97 PheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAsp 116  
 Db 241 TTTTCTTTTGAAGACTACTGATTTTACACTGATGATGATGATGATGATGATGATGATGATG 300  
 Qy 117 LeuGlnProGlyIleValLeuTyrProThrValLeuPheIleTyrGlyGlyProGln 134  
 Db 301 CTACAACTCGAAAGAAATACCCCACTGTGTATTCATATATATATATATATATATATATAT 360  
 Qy 134 134  
 Db 361 CTGGTGAACAATCGGTTTAAAGAGTCAAGTATTTCCGCTGACACCCCTGCGCTCCCTG 420  
 Qy 134 134  
 Db 421 GGTATATGTGTGTGTGTATACAAACAGGGATCTGTCAACGAGACTTAATTTGAA 480  
 Qy 135 135  
 Db 481 GGCCTCTTAAAT 540  
 Qy 148 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPhe 167

Db 541 TACCTAGACATCTAGATGACTTCTTATGATCTTGATTCGATGCGGATCCACGCTGTCC 600  
 Qy 168 TyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAla 187  
 Db 601 TATGTGTGTACTCTCTCTCTGATGACATTAAAGCAGAGGTGCAATATCTTCCGGGTGC 660  
 Qy 187 AlaLeuAlaAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyr 207  
 Db 661 TATTCTGGGGCCCAAGTACACCTCTGTGATCTTCTATATATACAGATACACGAGGCTTA 720  
 Qy 207 TWTGlyHisProAspGln 213  
 Db 721 TATGGTCAACCTGACACAG 739

RESULT 11  
 B0675260 910 bp mRNA linear EST 15-JUL-2002  
 LOCUS AGENCOURT 8354972 NIH\_MGC\_102 Homo sapiens cDNA clone IMAGE:6275478  
 DEFINITION 5', mRNA sequence.  
 ACCESSION B0675260  
 VERSION B0675260.1 GI:21786094  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 910)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strusberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 o DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:  
<http://image.llnl.gov>  
 Plate: LTCM2458 row: 1 column: 07  
 High quality sequence stop: 618.

#### FEATURES

source 1..910  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6275478"  
 /tissue\_type="epidermoid carcinoma, cell line"  
 /lab\_host="RDH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_102"  
 /note="Organ: salivary gland; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAGG). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and SuperScript II RT (Life Technologies). Note: this is a NIH MGC Library."  
 BASE COUNT 248 a 186 c 212 g 262 t 2 others  
 ORIGIN

#### Alignment Scores:

Pred. No.: 5 336-101 Length: 910  
 Score: 991.00 Matches: 195  
 Percent Similarity: 77.25% Conservative: 2  
 Best Local Similarity: 76.47% Mismatches: 53  
 Query Match: 58.99% Indels: 53  
 DB: 13 Gaps: 2

US-10-070-464-3 (1-310) x B0675260 (1-910)

Qy 1 PheGluGlyThrIleAspSerProGluIleHisIleLeuTyrValIleSerTyrValAsn 20



Db 605 CATGGCTTCCTGATGAGATGTCATTTTGCACATACCAATATATTACGAGTTT 664  
 Qy 260 euVal-ArgAlaGly-LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSer11 279  
 Db 665 TAGTTGAGGGCTGGAGAAAGCCATATATATTACGATCTATCCCTCAGAGACACAGCAT 724  
 Qy 279 eArGValProGluSerGlyGluHisTyrGluLeuHisTyrLeuGlnGluAsp 299  
 Db 725 MAGAGC-CCTGAATCGGAGAACTTATGAACTGCTCTTGCC--TACCTTCAGAAAA 780  
 Qy 299 nleu 300  
 Db 781 CCTG 784  
 RESULT 13  
 BU447277 735 bp mRNA linear EST 29-NOV-2002  
 LOCUS 603766685F1 CSEORBN13 Gallus gallus clone CHEST693j22 5', mRNA  
 DEFINITION sequence.  
 ACCESSION BU447277  
 VERSION BU447277.1 GI:25936588  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 735)  
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Boesch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22355534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology (UMIST)  
 )  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
 Location/Qualifiers  
 FEATURES  
 source 1..735  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Layer"  
 /db\_xref="taxon:9031"  
 /clone="CHEST693j22"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEORBN13"  
 /note="Organ: ovary; Vector: pBluescript II KS(+); Site\_1:  
 EcoRI; Site\_2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was bluntended, ligated to NotI adapters, digested with EcoRI  
 , size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

BASE COUNT 183 a 170 c 186 g 196 t  
 ORIGIN  
 Alignment Scores: 8.17e-98 Length: 735  
 Pred. No.:

Score: 961.50 Matches: 186  
 Percent Similarity: 78.694 Conservative: 6  
 Best Local Similarity: 76.234 Mismatches: 2  
 Query Match: 57.234 Indels: 50  
 Db: 13 Gaps: 1  
 US-10-070-464-3 (1-310) x BU447277 (1-735)  
 Qy 110 MelLeuTyrLysProHisAspLeuGlnProGlyLysValTyrProThrValLeuPhe11e 129  
 Db 6 ATGATGTACAAACCTTCACATCTGCACAACTGGAAAAAGATACCTTACGTGATCTTCATC 65  
 Qy 130 TyrGlyGlyProGln----- 134  
 Db 66 TATGAGGCCCTCAGGTGACGCTAGTGAACAATTCGATTTAAAGACTCAAAATTTCCGA 125  
 Qy 134 ----- 134  
 Db 126 TTGAACACCTTGACCTCTTAGGCTATGTTGTTGTTATTGACAACCGGGGCTCTGC 185  
 Qy 135 -----GlyGlnIleGluLeuAsp 140  
 Db 186 CACCGAGGCTGAAAGTTTGAAAGAGCCTTTAAATACAAATAGGACAAATAGAAATTGAT 245  
 Qy 161 AapGlnValGluGlyLeuGlnTyrLeuAlaSerGlyTyrAspPhe11eAspLeuAspArg 160  
 Db 246 GACCAAGTGAAGAGCTGCACTACTTAGGCTCCGCTAGTACTTCTTATTTGGATCGT 305  
 Qy 161 ValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180  
 Db 306 GTTGGCATTCAGAGGCTGGTCCATAGGAGCTACCTCTCTTATGSCCTTTAATGACAGAG 365  
 Qy 181 SerAsp11ePheArgValAlaIleAlaGlyAlaProValThrLeuTrp11ePheTyrAsp 200  
 Db 366 TCAGATATCTTCAGAGGTTGCGATCGCGGAGACCCGTCACCCCTGGAGATTTTCTATGAC 425  
 Qy 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeu 220  
 Db 426 ACGGGTACACGAGAGCGCTACATGAGCCACCGGACACAGAGAGGAGGCTATTAACCG 485  
 Qy 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArg-LeuLeuLeuLe 240  
 Db 486 GGTTCAGTGGCCATCGACAGCTAGAAAGTTCTCTTGAAACCAACCGTTTGTGCTGGCT 545  
 Qy 240 uHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSer11eLeuLeuSerPheLe 260  
 Db 546 ACATGGGTTCTTGGATGAGAAAGTTCACCTTGCACACACTAGTATTTTGCTCAGCTTTT 605  
 Qy 260 uValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSer11eAr 280  
 Db 606 AGTGAGAGCTGGAGAAAGCCATACGACTTGCAGATCTACCTCAGAGAGGACACAGTATAG 665  
 Qy 280 gValProGluSerGlyGluHisTyrGluLeuHisTyrLeuGlnGluAsnLe 300  
 Db 666 GGTGCTGAGTGGGAGAGCACTATGAACTCATCTACGTATTTACCTGCAAGAGATCT 725  
 Qy 300 uGlySerArg 303  
 Db 726 GGGCTCTCGC 735  
 RESULT 14  
 BI084885/c 855 bp mRNA linear EST 20-JUN-2001  
 LOCUS 602869453T1 NIH\_MGC\_102 Homo sapiens cDNA clone IMAGE:5013996 3',  
 DEFINITION mRNA sequence.  
 ACCESSION BI084885  
 VERSION BI084885.1 GI:14503215  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 855)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: CLONETECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LCM1820 row: j column: 13  
 High quality sequence start: 3  
 High quality sequence stop: 844.  
 Location/Qualifiers

## FEATURES

source

1..855  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5013996"  
 /issue\_type="epidermoid carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH MGC 102"  
 /note="Torgan: salivary gland; Vector: pOT87; Site\_1: XhoI;  
 Site\_2: EcoRI; CDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCAAGG(G). Library constructed  
 by Ling Hong in the laboratory of Gerald M. Rubin  
 (University of California, Berkeley) using ZAP-CDNA  
 synthesis kit (Stratagene) and Superscript II RT (Life  
 Technologies). Note: this is a NIH-MGC Library."  
 BASE COUNT 244 a 200 c 172 g 239 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 3,14e-95 Length: 855  
 Score: 939.50 Matches: 196  
 Percent Similarity: 77.738 Conservative: 3  
 Best Local Similarity: 76.568 Mismatches: 3  
 Query Match: 55.924 Indels: 55  
 DB: 12 Gaps: 1

US-10-070-464-3 (1-310) x BI084885 (1-855)

QY 109 GlyMetLeuTyrLysProHisAsp-LeuGlnPro-GlyLysLysTyrProThrValLeuP 128  
 DB 854 GGGATGCTCTACAGGCTCAGAGCTCTACAGCTGGGAAAGATATCTTACTGTGCTGT 795  
 QY 128 heileTyrGlyGlyProGln----- 134  
 DB 794 TCATTAATGATGTCCTCTCA-GGTGCAATTGGTGAATATCGGTTTAAAGAGTCAAGTAT 736  
 QY 134 ----- 134  
 DB 735 TTCGGTTGAATACCTAGCCTCTAGGTTATGCGTTGTAGTATAGACAACAGGGA 676  
 QY 135 -----GlyGlnIleG 138  
 DB 675 TCCTGTCACCGAGGGCTTAATATTGAAGCGCCTTAATATAAATAGGTCAAATAG 616  
 QY 138 LuLeAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAsp-PheIleAsp 157  
 DB 615 AAATTGACATCAGTGGAGAGACTCCATATCTAGCTTCTGATATGAGTTTCATGAC 556  
 QY 158 LeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeu 177  
 DB 555 TTAGATCGTGTGGGATCCACGGCTGCTCTATGAGAGATACCTCTCCCTGATGGCATTA 496  
 QY 178 -MetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrI 197  
 DB 495 ACTGACAGAGGTCAGATATCTTCAGGGGTGCTATGCTGGGCCCCAGTCACTCTGTGAT 436

QY 190 ePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnG 217  
 DB 435 CTTTATGATATACAGATACACCGAAAGTATATGGTTCACCTGCACCAATGAAACAGCG 376  
 QY 217 YTTTyrLeuGlySerValAlaMetGlnAlaIleLysPheProSerGluProAsnArgLe 237  
 DB 376 CTATTACTTAAGATCTGTGGCCATGCAAGCAAAAAAGTTCCCTCGAACCAAAATGCTT 316  
 QY 237 UdeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 257  
 DB 315 ACTGCTCTTACATGGTTCTCTGATGAGATATCCATTTTGACATACAGATATATTA 256  
 QY 257 USerPheLeuValAlaGlyAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgH 277  
 DB 255 GAGTTTATAGAGAGGCTGGAAGCAATATGATTAACATCTATCTCCAGAGAGACA 196  
 QY 277 aSerIleArgValAlaProGluSerGlyGlnHisTyrGluLeuHisLeuLeuHisTyrLeuL 297  
 DB 195 CAGCATTAAGAGTTCTCGAATCGGAGAACATTATGAACGATCTTTTGACACTTCA 136  
 QY 297 nGluAsnLeuGlySerArgIleAlaAlaLeuIleValIle 310  
 DB 135 AGAAAACTTGATCACGATATGCTGCTCTAAAGTGATA 96

RESULT 15  
 AL582206/ 1085 bp mRNA linear EST 01-JUN-2003  
 LOCUS AL582206 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 DEFINITION Homo sapiens CDNA clone CS0DL005YD02 3-PRIME, mRNA sequence.  
 ACCESSION AL582206  
 VERSION AL582206.2 GI:31320424  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 TITLE Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 JOURNAL Full-length CDNA libraries and normalization  
 COMMENT Unpublished  
 On Feb 16, 2001 this sequence version replaced gi:12949961.

CONTACT: Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 7542.r for  
 more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DL005YD02&INP1cluster=7542.r>. Contact :  
 Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>  
 Faraday Avenue Genoscope sequence ID : CS0DL005YD01NP1.

## FEATURES

source

1..1085  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL005YD02"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_id="RAMOS CELL LINE"  
 /note="1st strand CDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand CDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## BASE COUNT

297 a 221 c 220 g 315 t 32 others

## ORIGIN

Alignment Scores:  
 Pred. No.: 6.72e-95 Length: 1085  
 Score: 938.00 Matches: 221

```
Percent Similarity: 70.40% Conservative: 5
Best Local Similarity: 68.85% Mismatches: 7
Query Match: 55.83% Indels: 2
DB: 9 Gaps: 5
```

### Best Local Similarity:

Query Match: 55.83%

**DB:**

Conservative: 5  
Mismatches: 2

**Mismatches:**

Indels: 72

**Gaps: 5**

US-10-070-464-3 (1-310) x AL582206 (1-1085)

Job time : 1876.46 secs

QY	2	GIuGIuYThrLysAspSerProLeuGluHisLLeuTYrValValSerTYrValAspPro	21
Db	1021	GAAGGACCAACAAATCCCTTTCAGAGMATCATCTTACGTACGMAATTACGTAAATCT	962
QY	22	GIuGIuValThrArgLeuThrAspArgGlyTYrSerHisSerCysLLeuSerGlnHis	41
Db	961	GGAGAGGTGAMAAAGCTGACTGACCGGCTACTCATCTTGTGCTGCATCAGTCAGCAC	902
QY	42	CysAspPhePheLLeuSerLYrTYrSerAsnGlnLysAsnProHisCysValSerLeuTYr	61
Db	901	TGTGACTCTTTTAATAAKTAAGTAATAGTAATCAACAAAGAAATTCACACTGTGTGCTCTTAC	842
QY	62	LysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIle	81
Db	841	AAGCTATCAAGTCCTGAAGATGACCCCACTTGCAAAACAAAGAAATTTTGGCTACCAT	782
QY	82	LeuAspSerValGlyProLeuProAspTYrThrProProGluLLeuPheSerPheGluSer	101
Db	781	TTCGATTTCAGCAGGCTCTCTCTGCTGACTATACCTCCAGAAATTTTCTCTTTGAAAGT	722
QY	102	ThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGlnProGlyLys	121
Db	721	ACTACTGGATTACATATGATGGAGGAGCTCTCAACAGCTCATATCATCTACAGCTGGAAAG	662
QY	122	LysTYr-ProThrValLeuPheLLeu-TyrGlyGly-ProGlnGlyGlnIleGluLLeuAsp	140
Db	661	ATATATCKCTACTGCTGCTCTTCATATTAAGTGATGCTGCTGACAGTAAATA-----	611
QY	141	AspGlnValGluGlyLeuGlnTYrLeuAlaSerArgTYrAspPheLLeuAspAspArg	160
Db	611	-----	611
QY	161	ValGlyLLeuHisGlyTYrTrpSerTYrGlyGlyTYrLeuSerLeuMetAlaLeuMetGlnArg	180
Db	611	-----	611
QY	181	SerAspLLeuPheArgValAlaIleAlaGlyAla-ProValThr-LeuTrpLLeuPheTYr-	199
Db	610	-----TTTGCTGGGGGGCCCCAGTACACTCTGTGGATCATTTACTG	572
QY	200	--AspThrGlyTYrThrGlnArg-TYrMetGly-HisProAspGlnAsn-GluGlnGly-	217
Db	571	ATGATATACAGGATATACACGGAACGATTAATATGGCTCACCTGTACACAGATATGAACAGGAC	512
QY	218	TYrTYrLeuGly-SerValAlaMetGlnAlaGluLys-PheProSerGluProAsnArgL	237
Db	511	TATTACTTAGAGTGGCTGTGGCCATGCAGAACAGAAATAGTAAACCCCTTGAAACCAATCGTT	452
QY	237	euleLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAlaHisTrpSerLLeuL	257
Db	451	TACTCTCTTAATAGGTTTCCGTGAGTGAAGAAATGTCATTTTGACACATACAGTATATTAC	392
QY	257	eusErPheLeuValArg--AlaGlyLysProTYr--AspLeuGlnLLeuTYrProGlnG	275
Db	391	TGAGTTTTTTAGTCCAGTCCGCTGGAACGC-CATATGTCATTTA-----CAGG	345
QY	275	luArgHisSerLLeuArgValProGluSerGlyGluHisTYrGluLeuHisLLeuLeuHis-	294
Db	344	AGAGACACAGCATTAAGAGTTCTCGAATCGGAGAAACATATGAACCTGTCCTTTGCACT	285
QY	295	-TYrLeuGlnGlnuLeuLeuGlySerArgLLeuAlaAla	306
Db	284	GACTTTCAGAGAAACCTATGGATGCACGATATTAGCT	248

Search completed: October 16, 2003, 03:25:10

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 15:44:49 ; Search time 28.3565 Seconds  
(without alignments)  
2602.858 Million cell updates/sec

Title: US-10-070-464-5

Perfect score: 2482

Sequence: 1 TGTANPKVTFKMSEIMIDAE.....HLHLVQLGNLGRIRALKVY 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2482	100.0	465	22	Human DPP8 318Thr-
2	2482	100.0	724	23	ABB97362
3	2482	100.0	782	23	ABB97361
4	2441.5	98.4	632	22	Human protein sequ
5	2442	97.6	882	22	Human DPP8. Homo
6	2442	97.6	882	23	AAE24170
7	2442	97.6	882	23	Human DPP4 related
8	2442	97.6	882	23	Human protease PRT
9	2442	97.6	882	23	AAU74749
					Amino acid sequenc

10	2422	97.6	882	24	ABU07720	Human serine prote
11	2244.5	90.4	580	23	AAE14337	Human protease PRT
12	1809.5	72.9	690	23	ABG61594	Human DPP-1 splic
13	1808	72.8	658	23	ABG61600	Human DPP-1 splic
14	1808	72.8	661	23	ABG61596	Human DPP-1 splic
15	1540	62.0	613	23	ABG61601	Human DPP-1 splic
16	1494.5	60.2	830	23	AAE24171	Human dipeptidyl p
17	1494.5	60.2	863	23	ABG61592	Human DPP4 relate
18	1494.5	60.2	882	23	ABG61602	Human DPP4 relate
19	1494.5	60.2	882	23	ABG61602	Human DPP4 relate
20	1494.5	60.2	882	23	ABG61604	Human DPP4 relate
21	1494.5	60.2	882	23	AAE24168	Human dipeptidyl p
22	1488.5	60.0	869	23	AAE24169	Human dipeptidyl p
23	1445	58.2	879	23	ABG61607	Human DPP-2 splic
24	1445	58.2	879	23	ABG61608	Human DPP-2 splic
25	1377.5	55.5	310	22	ABG61608	Human DPP8 524phe-
26	1377.5	55.5	310	22	ABG61608	Human dipeptidyl p
27	1369	55.2	847	23	AAE23875	Human dipeptidyl p
28	1298.5	52.3	497	23	ABG64845	Human albumin fusi
29	1298.5	52.3	497	23	AAU96169	Human secreted pro
30	1296	52.2	683	22	AAU40510	Human polypeptide
31	1296	52.2	737	22	AAU38724	Human polypeptide
32	1294.5	52.2	497	23	ABG64844	Human albumin fusi
33	1294.5	52.2	497	23	AAU96192	Human secreted pro
34	1187.5	47.8	508	21	AAE42928	Human ORFX ORF2692
35	1094.5	44.1	832	23	ABG61605	Human DPP-2 splic
36	1094.5	44.1	832	23	ABG61606	Human DPP-2 splic
37	1045	42.1	819	23	ABG61609	Human DPP-2 splic
38	1045	42.1	819	23	ABG61610	Human DPP-2 splic
39	1008.5	40.6	720	21	AAE41626	Human ORFX ORF1390
40	860.5	34.7	360	22	AAE47190	Human DPP8 244Glu-
41	845	34.0	1042	22	ABG60137	Drosophila melanog
42	845	34.0	1102	22	ABG62029	Drosophila melanog
43	744	30.0	250	23	ABG99949	Dipeptidyl peptida
44	663	26.7	136	22	AAE74673	Human protease and
45	663	26.7	193	23	ABE89739	Human polypeptide

#### ALIGNMENTS

```

RESULT 1
AAB47189
ID AAB47189 standard; Protein; 465 AA.
AC AAB47189;
DT 29-JUN-2001 (first entry)
XX
DE Human DPP8 318Thr-658Val-759Ala-882Ile.
XX
KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW dipeptidyl peptidase; DPP4; T cell; cleavage; diarrhoea;
KW growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus; glucose intolerance;
KW immunosuppression.
XX
OS Homo sapiens.
XX
PN WO200119866-A1.
XX
PD 22-MAR-2001.
XX
PF 11-SEP-2000; 2000MO-AU01085.
XX
PR 10-SEP-1999; 99AU-0002762.
PR 18-FEB-2000; 2000AU-0005709.
XX
PA (UNSW ) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI; 2001-281520/29.

```



DR N-PSDB; AAC85696.  
XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving  
PT substrates, identifying inhibitors of DPP8 catalytic activity which  
PT have therapeutic uses, and for detecting activated T cells -  
XX  
XX Claim 5; Page 71-72; 78pp; English.  
XX  
XX The sequences given in AAB47188-90 represent fragments of human  
CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for  
CC H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a  
CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable  
CC of hydrolysing the peptide bond C-terminal to Pro in each of these  
CC compounds. DPP8 is homologous with human DPP4V. DPP8 is useful for  
CC cleaving a substrate, and for detecting an activated T cell which  
CC involves measuring the level of DPP8 gene expression in a T cell. The  
CC level of DPP8 expression is detected by detecting the amount of DPP8  
CC RNA in the cell. It is also useful for identifying a molecule capable  
CC of inhibiting the cleavage of the substrate by DPP8. Molecules  
CC identified as inhibiting DPP8 catalytic activity may be useful for  
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels  
CC in non-insulin dependent diabetes mellitus and other disorders  
CC involving glucose intolerance, enhancing mucosal regeneration and  
CC as immunosuppressants.  
XX  
XX

SQ Sequence 465 AA;

Query Match 100.0%; Score 2482; DB 22; Length 465;

Best Local Similarity 100.0%; Pred. No. 6.4e-237;

Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPFELFEGVEYIARAGWPBGKYAMSI 60  
DB 1 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPFELFEGVEYIARAGWPBGKYAMSI 60  
QY 61 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 120  
DB 61 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 120  
QY 121 HVPFQSHHEEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDPKCIKEIA 180  
DB 121 HVPFQSHHEEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDPKCIKEIA 180  
QY 121 HVPFQSHHEEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDPKCIKEIA 180  
DB 121 HVPFQSHHEEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDPKCIKEIA 180  
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHNLVYVYVNGEYTRLLDRGYSH 240  
DB 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHNLVYVYVNGEYTRLLDRGYSH 240  
QY 241 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPCKTKEFWATILLDSAGPLPDTTP 300  
DB 241 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPCKTKEFWATILLDSAGPLPDTTP 300  
QY 301 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPOVAIAGAPVTLMIYFDGYTE 360  
DB 301 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPOVAIAGAPVTLMIYFDGYTE 360  
QY 361 RYWGHPDQNEQGYVLSVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILLSLVRAK 420  
DB 361 RYWGHPDQNEQGYVLSVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILLSLVRAK 420  
QY 421 PYLOLQIYPOBRHSIRVPSGHEHYELHLHYLOENLGSRIAAKVT 465  
DB 421 PYLOLQIYPOBRHSIRVPSGHEHYELHLHYLOENLGSRIAAKVT 465

RESULT 2

ID ABB97362 standard; Protein; 724 AA.

AC ABB97362;

XX 27-JUN-2002 (first entry)

XX Novel human protein SEQ ID NO: 630.

XX Human; Antianemic; vulnerary; antiinflammatory; immunomodulator;  
KM antiinfectility; cerebroprotective; cytosarctic; rheumatic; gene therapy;  
KM neuroprotective; antiparkinsonian; protein therapy; EST;  
KM expressed sequence tag.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200222660-A2.  
XX  
XX 21-MAR-2002.  
XX  
XX 10-SEP-2001; 2001WO-US26015.  
XX  
XX 11-SEP-2000; 2000US-0659671.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Auendi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Dirmnac RT;  
XX WPI: 2002-292408/33.  
DR N-PSDB; ABR32548.  
DR  
XX  
XX

An isolated polynucleotide for treating diseases associated with its encoded polypeptide such as cancer and multiple sclerosis -

Example 2; SEQ ID NO 630; 509pp; English.

CC The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention.  
XX  
XX

SQ Sequence 724 AA;

Query Match 100.0%; Score 2482; DB 23; Length 724;

Best Local Similarity 100.0%; Pred. No. 1.3e-236;

Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPFELFEGVEYIARAGWPBGKYAMSI 60  
DB 260 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPFELFEGVEYIARAGWPBGKYAMSI 319  
QY 61 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 120  
DB 320 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 379  
QY 121 HVPFQSHHEEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDPKCIKEIA 180  
DB 380 HVPFQSHHEEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDPKCIKEIA 439  
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHNLVYVYVNGEYTRLLDRGYSH 240  
DB 440 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHNLVYVYVNGEYTRLLDRGYSH 499  
QY 241 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPCKTKEFWATILLDSAGPLPDTTP 300  
DB 500 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPCKTKEFWATILLDSAGPLPDTTP 559  
QY 301 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPOVAIAGAPVTLMIYFDGYTE 360  
DB 560 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPOVAIAGAPVTLMIYFDGYTE 619  
QY 361 RYWGHPDQNEQGYVLSVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILLSLVRAK 420  
DB 620 RYWGHPDQNEQGYVLSVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILLSLVRAK 679

QY 421 PYDQIYPERHSIRVPESEGEHYELHLHYLOENLSRIAALVKYI 465  
 DB 680 PYDQIYPERHSIRVPESEGEHYELHLHYLOENLSRIAALVKYI 724

RESULT 3  
 ABB97361 standard; Protein; 782 AA.

AC ABB97361;

DT 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 629.

KM Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;  
 KW antifertility; cerebroprotective; cytosolic; rheumatic; gene therapy;  
 KM neuroprotective; antiparkinsonian; protein therapy; EST;  
 XX expressed sequence tag.

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wenman T, Drmanac RT;

DR WPI; 2002-292408/33.

DR N-PSDB; ABN32547.

PT An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis -

PS Example 2; SEQ ID NO 629; 509pp; English.

CC The present invention provides the protein and coding sequences of 444  
 CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to help tissue regrowth  
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a protein of the invention.

SQ Sequence 782 AA;

Query Match 100.0%; Score 2482; DB 23; Length 782;

Best Local Similarity 100.0%; Pred. No. 1.5e-236;

Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKTFKMSIEMIDAGRIIDVIDELIOPFIEFGVYIARAGTPEGKAWMSI 60  
 DB 318 TGTANPKTFKMSIEMIDAGRIIDVIDELIOPFIEFGVYIARAGTPEGKAWMSI 377  
 QY 61 LIDRSQTRQIVLISSELPFVDDVMEORQLIESVDSVTPLIYEETDININIDIF 120  
 DB 378 LIDRSQTRQIVLISSELPFVDDVMEORQLIESVDSVTPLIYEETDININIDIF 437  
 QY 121 HVFPSHEIEIFIPASECKTGRHLKYITSLIKESKYSRSGGLPAPSDFCPIKEEA 180  
 DB 438 HVFPSHEIEIFIPASECKTGRHLKYITSLIKESKYSRSGGLPAPSDFCPIKEEA 497

QY 181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVSVNGEYTRLTDRGYSH 240  
 DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVSVNGEYTRLTDRGYSH 557  
 QY 241 SCCISQHCDFEISKYSNOKNPHCVSLYKLSPPEDDPTCKTKEFWATILDSAGPLPDYTPP 300  
 DB 558 SCCISQHCDFEISKYSNOKNPHCVSLYKLSPPEDDPTCKTKEFWATILDSAGPLPDYTPP 617  
 QY 301 EIFSFEETGFTLYGMLYKPHDLQPGKYPPTYVFIYGGPQVALAGAPVTLMIFPDGYTE 360  
 DB 618 EIFSFEETGFTLYGMLYKPHDLQPGKYPPTYVFIYGGPQVALAGAPVTLMIFPDGYTE 677  
 QY 361 RYMGHPDQNEQGYLLASVMAQAEKFPSEPNRLLLGLFDENYHFAHTSILSFLVRAGK 420  
 DB 678 RYMGHPDQNEQGYLLASVMAQAEKFPSEPNRLLLGLFDENYHFAHTSILSFLVRAGK 737  
 QY 421 PYDQIYPERHSIRVPESEGEHYELHLHYLOENLSRIAALVKYI 465  
 DB 738 PYDQIYPERHSIRVPESEGEHYELHLHYLOENLSRIAALVKYI 782

RESULT 4  
 AAB93565 standard; Protein; 632 AA.

AC AAB93565;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12964.

KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -

PS Claim 8; SEQ ID 12964; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13632 to AAH18742 represent human cDNA sequences; AAB9246 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 632 AA;

Query Match 98.4%; Score 2441.5; DB 22; Length 632;  
 Best Local Similarity 90.3%; Pred. No. 1.1e-232;  
 Matches 464; Conservative 0; Mismatches 1; Indels 49; Gaps 1;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 60  
 DB 119 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 178  
 QY 61 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETDIDMINIHDF 120  
 DB 179 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETDIDMINIHDF 238  
 QY 121 HVPQSHHEEIEIFAFSECKTGFRHLKYKTSILKESKYKSSGGLPAPSDPKPIKEIA 180  
 DB 239 HVPQSHHEEIEIFAFSECKTGFRHLKYKTSILKESKYKSSGGLPAPSDPKPIKEIA 298  
 QY 181 ITSGEWELVGRHGSNIQVDEVRRLVFEETKQSPLEHNLVYVSVNGEYTRLTDRGYSH 240  
 DB 299 ITSGEWELVGRHGSNIQVDEVRRLVFEETKQSPLEHNLVYVSVNGEYTRLTDRGYSH 358  
 QY 241 SCCISQCHDFIFISKYSNQKAPHCVSILYKLSSPEDDPTCKTEFWATILDSAGLPDYTPP 300  
 DB 359 SCCISQCHDFIFISKYSNQKAPHCVSILYKLSSPEDDPTCKTEFWATILDSAGLPDYTPP 418  
 QY 301 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPQ----- 340  
 DB 419 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPQVQVLNNRFKGVKRYRLNTLA 478  
 QY 341 -----VAIAGAPTLWIFPDGTGYTERYMGHPONEQ 371  
 DB 479 SLGIVVVVINDRGSSCHRGKLEBGAFFKXKVAIAGAPTLWIFPDGTGYTERYMGHPONEQ 538  
 QY 372 GYLLGSVVAQAEKFPSEPNRLLLHGFLDENVHFAHTSILSFLVRAKRPYDLOIYPOER 431  
 DB 539 GYLLGSVVAQAEKFPSEPNRLLLHGFLDENVHFAHTSILSFLVRAKRPYDLOIYPOER 598  
 QY 432 HSIRVPESGEHYELHLHYIQENLGSRIALAKVI 465  
 DB 599 HSIRVPESGEHYELHLHYIQENLGSRIALAKVI 632

RESULT 5  
 AAB47187  
 ID AAB47187 standard; Protein; 882 AA.

AC AAB47187;

DT 29-JUN-2001 (first entry)

XX Human DPP8.

XX Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;

KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;

KW growth hormone deficiency; glucose level; mucosal regeneration;

KW non-insulin dependent diabetes mellitus; glucose intolerance;

XX immunosuppression.

XX Homo sapiens.

OS

XX

Key Location/Qualifiers

FT Active-site 739  
 FT /note= "Forms part of Ser-Asp-His catalytic triad"  
 FT Active-site 817  
 FT /note= "Forms part of Ser-Asp-His catalytic triad"  
 FT Active-site 839  
 FT /note= "Forms part of Ser-Asp-His catalytic triad"  
 PN MO200119866-A1.  
 XX  
 PD 22-MAR-2001.  
 XX  
 XX 11-SEP-2000; 2000MO-AU01085.  
 XX  
 XX 10-SEP-1999; 99AU-0002762.  
 PR 18-FEB-2000; 2000AU-0005709.  
 XX  
 XX (UNSY ) UNIV SYDNEY.  
 PA  
 XX Abbot CA, Gorell MD;  
 PI  
 XX WPI; 2001-281520/29.  
 DR N-PDB; AAC85694.  
 XX  
 PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving  
 PT substrates, identifying inhibitors of DPP8 catalytic activity which  
 PT have therapeutic uses, and for detecting activated T cells  
 XX  
 PS Claim 1; Fig 2; 78pp; English.

CC This sequence represents human dipeptidyl aminopeptidase (DPP8).  
 CC DPP8 has substrate specificity for H-Ala-pro-pNA, H-Gly-pro-pNA and  
 CC H-Ary-pro-pNA. Therefore, it is a prolyl oligopeptidase and a  
 CC dipeptidyl peptidase, because it is capable of hydrolysing the  
 CC peptide bond C-terminal to pro in each of these compounds. DPP8  
 CC is homologous with human DPPIV. DPP8 is useful for cleaving a  
 CC substrate, and for detecting an activated T cell which involves  
 CC measuring the level of DPP8 gene expression in a T cell. The level  
 CC of DPP8 expression is detected by detecting the amount of DPP8 RNA  
 CC in the cell. It is also useful for identifying a molecule capable  
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules  
 CC identified as inhibiting DPP8 catalytic activity may be useful for  
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels  
 CC in non-insulin dependent diabetes mellitus and other disorders  
 CC involving glucose intolerance, enhancing mucosal regeneration and  
 CC as immunosuppressants.

XX Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 22; Length 882;  
 Best Local Similarity 82.3%; Pred. No. 1.6e-230;  
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 60  
 DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 377  
 QY 61 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETDIDMINIHDF 120  
 DB 378 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETDIDMINIHDF 437  
 QY 121 HVPQSHHEEIEIFAFSECKTGFRHLKYKTSILKESKYKSSGGLPAPSDPKPIKEIA 180  
 DB 438 HVPQSHHEEIEIFAFSECKTGFRHLKYKTSILKESKYKSSGGLPAPSDPKPIKEIA 497  
 QY 181 ITSGEWELVGRHGSNIQVDEVRRLVFEETKQSPLEHNLVYVSVNGEYTRLTDRGYSH 240  
 DB 498 ITSGEWELVGRHGSNIQVDEVRRLVFEETKQSPLEHNLVYVSVNGEYTRLTDRGYSH 557  
 QY 241 SCCISQCHDFIFISKYSNQKAPHCVSILYKLSSPEDDPTCKTEFWATILDSAGLPDYTPP 300  
 DB 558 SCCISQCHDFIFISKYSNQKAPHCVSILYKLSSPEDDPTCKTEFWATILDSAGLPDYTPP 617  
 QY 301 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPQ----- 340

```

Db      618 EIFFESTTGTFLYGMLYKPHDLPQKKPTVTLFYGGPQVQVNNRFGVKYFRLLNTLA 677
      341 -----
Qy      678 SLGYVVVVVINDRNGSCHRLKFEKAFKYGKQIEIDDQVEGLQYLASRYDFILDRVGJHG 737
Db      341 -----
Qy      738 WSYGYLSLMLMQRSDIFRVALIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLYGSVAM 797
Db      381 QAEKFPSEPNRLLHGFLENVHFAPHTSILSFVLRAGKPYDLOIYQERHSIRVPESG 440
      798 QAEKFPSEPNRLLHGFLENVHFAPHTSILSFVLRAGKPYDLOIYQERHSIRVPESG 857
Qy      441 EHYELHLHYLOENLGSRIAALKVI 465
      858 EHYELHLHYLOENLGSRIAALKVI 882

```

## RESULT 6

AAE24170 ID AAE24170 standard; Protein; 882 AA.

AAE24170;

23-SEP-2002 (first entry)

Human dipeptidyl peptidase 8 (DPP8) protein.

Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;  
 autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;  
 graft rejection; antidiabetic; antinflammatory; immunosuppressive;  
 antiviral; enzyme.

Homo sapiens.

WO200234900-A1.

02-MAY-2002.

29-OCT-2001; 2001WO-AU01388.

27-OCT-2000; 2000AU-0001076.

(UNSY) UNIV SYDNEY.

Abbot CA, Correll MD;

WPI; 2002-454646/48.

N-PSDB; AAD38956.

New dipeptidyl peptidase (DPP) peptidases, useful for screening

inhibitors of DPP catalytic activity, which may be employed to treat

e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft

rejection and HIV infection -

Example; Fig 1; 91pp; English.

The present invention relates to dipeptidyl peptidase (DPP) proteins and

polynucleotides encoding such proteins. The DPP peptidases are useful

for screening inhibitors of DPP catalytic activity. The inhibitors are useful

for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft

rejection and HIV (human immuno deficiency virus) infection. The present

Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 23; Length 882;

Best Local Similarity 82.3%; Pred. No. 1.6e-230;

Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

1 TGTAPKVTFKMSEIMDAEGRIIDVIDKELIQPEILFEGVEYIARAGWTFEGKYSAMI 60

```

Db      318 TGTAPKVTFKMSEIMDAEGRIIDVIDKELIQPEILFEGVEYIARAGWTFEGKYSAMI 377
      61 ILDRSQTRLOQIVLISPELFIPVEDDVMERORLIESVPSVTPLIYEETTDIMINIDIF 120
      378 ILDRSQTRLOQIVLISPELFIPVEDDVMERORLIESVPSVTPLIYEETTDIMINIDIF 437
Qy      121 HVPFQSHHEEIEFIPASECKTGFRHLKYITSLKSKYKRSAGGLPAPSDFCPIKEEA 180
      438 HVPFQSHHEEIEFIPASECKTGFRHLKYITSLKSKYKRSAGGLPAPSDFCPIKEEA 497
Db      181 ITSGMEVLYGRHGSNIQVDEVRLVYFEGTKDSPLEHLYVVSYPNPGVTLTRGVSH 240
      498 ITSGMEVLYGRHGSNIQVDEVRLVYFEGTKDSPLEHLYVVSYPNPGVTLTRGVSH 557
Qy      241 SCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFWATTILDSAGPLDPYTP 300
      558 SCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFWATTILDSAGPLDPYTP 617
Db      301 EIFFESTTGTFLYGMLYKPHDLPQKKPTVTLFYGGPQVQVNNRFGVKYFRLLNTLA 677
      618 EIFFESTTGTFLYGMLYKPHDLPQKKPTVTLFYGGPQVQVNNRFGVKYFRLLNTLA 677
Qy      341 -----
      678 SLGYVVVVVINDRNGSCHRLKFEKAFKYGKQIEIDDQVEGLQYLASRYDFILDRVGJHG 737
Db      341 -----
Qy      738 WSYGYLSLMLMQRSDIFRVALIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLYGSVAM 797
Db      381 QAEKFPSEPNRLLHGFLENVHFAPHTSILSFVLRAGKPYDLOIYQERHSIRVPESG 440
      798 QAEKFPSEPNRLLHGFLENVHFAPHTSILSFVLRAGKPYDLOIYQERHSIRVPESG 857
Qy      441 EHYELHLHYLOENLGSRIAALKVI 465
      858 EHYELHLHYLOENLGSRIAALKVI 882

```

## RESULT 7

ABG61591 ID ABG61591 standard; Protein; 882 AA.

ABG61591;

12-AUG-2002 (first entry)

Human DPPIV related serine protease DPP-1.

Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;  
 DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 dyskinesia; reproductive disorder; inflammatory disorder;  
 metabolic disorder.

Homo sapiens.

WO200231134-A2.

18-APR-2002.

12-OCT-2001; 2001WO-US31874.

12-OCT-2000; 2000US-240117P.

(FERR) FERRING BV.

Qi S, Akinsanya KO, Riviere PJ, Juntien J;

WPI; 2002-444176/47.

0

DR N-PSDB; ABK83322.  
 XX  
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
 PT viral infections, cancers, allergies, neurological disorders, or pain  
 PT  
 XX  
 PS Claim 17, Fig 1, 113pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 23; Length 882;  
 Best Local Similarity 82.3%; Pred. No. 1.6e-230;  
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 60  
 DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 377  
 QY 61 LLDRSQTRIQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIWINIHDF 120  
 DB 378 LLDRSQTRIQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIWINIHDF 437  
 QY 121 HVPFQSHHEEIEFIPESECKTGRHLKYITSIKESKYKSSGGLPAPSDFKPIKEEIA 180  
 DB 438 HVPFQSHHEEIEFIPESECKTGRHLKYITSIKESKYKSSGGLPAPSDFKPIKEEIA 497  
 QY 181 ITSGEWEVLGRHGSNIQVEVRRLVYFEGTKDPLHHLVYVYNPGEVTRLTIDGYSH 240  
 DB 498 ITSGEWEVLGRHGSNIQVEVRRLVYFEGTKDPLHHLVYVYNPGEVTRLTIDGYSH 557  
 QY 241 SCCISQCHDFISKYNOKNPHCVSLYKLSPPEDDPTCKTKEFWATILDSAGLPDYTPP 300  
 DB 558 SCCISQCHDFISKYNOKNPHCVSLYKLSPPEDDPTCKTKEFWATILDSAGLPDYTPP 617  
 QY 301 EIFSFESTTGFTLYGMLYKPHDLPQPKKYPTVLFIYGGPQ----- 340  
 DB 618 EIFSFESTTGFTLYGMLYKPHDLPQPKKYPTVLFIYGGPQVQLVNNRFKGVYFRNLTLA 677  
 QY 341 ----- 340  
 DB 678 SLGIYVVVINDNGSCHRGKLFBSGAFYKKGQIHDQVEGLQYLASRYDFIDLDVINGIH 737  
 QY 341 -----VALAGAPVTLMIFYDTGYTERYVGHGDQEGGYLGSVAM 380  
 DB 738 WSYGYGLSLMALMQRSDIFRVAIAGAPVTLMIFYDTGYTERYVGHGDQEGGYLGSVAM 797  
 QY 381 QAEKFPSEBRNLLHLGFLDENHFAHTSLLSFLVRACKPVDLQIYPOBRHSIRVPSG 440  
 DB 798 QAEKFPSEBRNLLHLGFLDENHFAHTSLLSFLVRACKPVDLQIYPOBRHSIRVPSG 857  
 QY 441 EHYEHLHLHYLOENLGSRIALAKVI 465  
 DB 858 EHYEHLHLHYLOENLGSRIALAKVI 882

RESULT 8  
 AAU74749

ID AAU74749 standard; Protein; 882 AA.  
 XX  
 AC AAU74749;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Human protease PRTS-9 protein sequence.  
 XX  
 KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;  
 KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;  
 KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;  
 KW cell proliferative disorder; developmental disorder; epilepsy;  
 KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;  
 KW reproductive disorder; endometriosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200198468-A2.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 13-JUN-2001; 2001MO-US19178.  
 XX  
 PR 16-JUN-2000; 2000US-212336P.  
 PR 22-JUN-2000; 2000US-213955P.  
 PR 29-JUN-2000; 2000US-215396P.  
 PR 07-JUL-2000; 2000US-216821P.  
 PR 14-JUL-2000; 2000US-218946P.  
 XX  
 PA (INCYTE GENOMICS INC.  
 XX  
 PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM,  
 PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA,  
 PI Wallia NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT,  
 PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L,  
 PI Kallik DA;  
 XX  
 DR MPI; 2002-090437/12.  
 DR N-PSDB; ABK12892.  
 XX  
 PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful  
 PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.  
 PT gastritis), cardiovascular (e.g. atherosclerosis) and cell  
 PT proliferative (e.g. cancer) disorders -  
 XX  
 PS Claim 1; Page 140-142; 177pp; English.  
 XX  
 CC The present invention relates to twenty one new human proteases,  
 CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and  
 CC polypeptides of the invention are useful in the diagnosis, treatment and  
 CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and  
 CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and  
 CC myocardial infarction, autoimmune/inflammatory e.g. acquired  
 CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell  
 CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker  
 CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.  
 CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and  
 CC endometriosis disorders. Numerous other examples of each disorder are  
 CC given in the specification. The present protein sequence represents  
 CC the human protease PRTS-9 protein of the invention.  
 XX  
 SQ Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 23; Length 882;  
 Best Local Similarity 82.3%; Pred. No. 1.6e-230;  
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 60  
 DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 377  
 QY 61 LLDRSQTRIQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIWINIHDF 120

```

Db      378 LDRSOTRLQIVLISPELFIPEDDVMEQRRLIESVDSVTPLIYEETTDIMINIDIF 437
Qy      121 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIKEEIA 180
Db      438 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIKEEIA 497
Qy      181 ITSGEWEVLGRHGSNIQVDEVRRLVFEETKOSPLEHMLYVSYVNPGEVTRLTDGYSH 240
Db      498 ITSGEWEVLGRHGSNIQVDEVRRLVFEETKOSPLEHMLYVSYVNPGEVTRLTDGYSH 557
Qy      241 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 300
Db      558 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 617
Qy      301 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQ----- 340
Db      618 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQVQVLNNRFGVKYFRNLNTLA 677
Qy      341 ----- 340
Db      678 SLGYYVVVVIDNRGSGRGLKFECAFKKYKMGQIETDQVEGLQYLARYPIDLDRVGING 737
Qy      341 -----VALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 380
Db      738 WSYGYLSLMLMQRSDFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 797
Qy      381 QAEKFPSEBNRLLLLHGFLENVHFHTSILLSFLVRACKPYDLQIYPOERSIRVPEEG 440
Db      798 QAEKFPSEBNRLLLLHGFLENVHFHTSILLSFLVRACKPYDLQIYPOERSIRVPEEG 857
Qy      441 EHYELHLHYLOENLGSRIAALKVI 465
Db      858 EHYELHLHYLOENLGSRIAALKVI 882

RESULT 9
AAG78415
ID      AAG78415 standard; Protein; 882 AA.
XX
AC      AAG78415;
XX
DT      12-APR-2002 (first entry)
XX
DE      Amino acid sequence of 21953 human prollyl oligopeptidase.
XX
KW      21953 prollyl oligopeptidase; antibody; proline; endopeptidase;
KW      cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW      neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW      antidiabetic; antiarthritic; antiaesthetic; antiinflammatory;
KW      diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW      Grave's disease; neuronal disorder; demyelinating disease.
XX
OS      Homo sapiens.
XX
PN      WO200179473-A2.
XX
PD      25-OCT-2001.
XX
PF      11-APR-2001; 2001WO-US40483.
XX
PR      18-APR-2000; 2000US-197508P.
XX
PA      (MILL-) MILLENNIUM PHARM INC.
XX
PI      Meyers RA, Williamson M;
XX
DR      MPI; 2002-034353/04.
XX
DR      N-PSDB; AAH99934.
XX
PT      New polypeptides 21953, member of human prollyl oligopeptidase family,
PT      useful as diagnostic targets and therapeutic agents for controlling
PT      cancer, lymphoma and leukemia
XX

```

```

PS      Claim 1; Page 102-103; 121pp; English.
XX
CC      This invention relates to an isolated 21953 human prollyl
CC      oligopeptidase, which is cytosolic, antidiabetic, antiarthritic,
CC      neuroprotective, antihypertensive, dermatological, antipsoriatic,
CC      antiepileptic, ophthalmological, antiinflammatory, nootropic,
CC      antiparasitomaian, anticonvulsant, gynaecological, vasorelaxant,
CC      antitumoral, cardiant, anticholesterolic, anorectic and
CC      metabolic in its action. Uses include gene therapy, expression or
CC      activity of 21953 protein modulator, it is useful for identifying a
CC      compound which binds to it and can be used in preventing, treating
CC      or detecting a cellular proliferative or differentiative disorder.
CC      The 21953 molecules can act as novel diagnostic targets and therapeutic
CC      agents for controlling disorders associated with the aberrant activity
CC      or degradation of peptide hormones e.g., disorders associated with cell
CC      differentiation and proliferation such as cancer, immune function,
CC      reproductive, neurological and cardiovascular function. The 21953
CC      molecules are thus useful for treating and preventing cellular
CC      proliferative and differentiative disorders, haematopoietic neoplastic
CC      disorders, immune disorders such as autoimmune diseases, diabetes
CC      mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC      neuronal disorders, demyelinating diseases, vascular disorders and
CC      metabolism or pain disorders. This sequence represents the amino
CC      acid sequence of 21953 human prollyl oligopeptidase.
XX
SQ      Sequence 882 AA:
Query Match 97.6%; Score 2422; DB 23; Length 882;
Best Local Similarity 82.3%; Pred. No. 1,6e-230;
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

Qy      1 TGTANPKYTFKMSIEMIDAEGRIDYIDKELLQPFILFEGCYFIARACKTPPEKAWMSI 60
Db      318 TGTANPKYTFKMSIEMIDAEGRIDYIDKELLQPFILFEGCYFIARACKTPPEKAWMSI 377
Qy      61 LDRSOTRLQIVLISPELFIPEDDVMEQRRLIESVDSVTPLIYEETTDIMINIDIF 120
Db      378 LDRSOTRLQIVLISPELFIPEDDVMEQRRLIESVDSVTPLIYEETTDIMINIDIF 437
Qy      121 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIKEEIA 180
Db      438 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIKEEIA 497
Qy      181 ITSGEWEVLGRHGSNIQVDEVRRLVFEETKOSPLEHMLYVSYVNPGEVTRLTDGYSH 240
Db      498 ITSGEWEVLGRHGSNIQVDEVRRLVFEETKOSPLEHMLYVSYVNPGEVTRLTDGYSH 557
Qy      241 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 300
Db      558 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 617
Qy      301 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQ----- 340
Db      618 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQVQVLNNRFGVKYFRNLNTLA 677
Qy      341 ----- 340
Db      678 SLGYYVVVVIDNRGSGRGLKFECAFKKYKMGQIETDQVEGLQYLARYPIDLDRVGING 737
Qy      341 -----VALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 380
Db      738 WSYGYLSLMLMQRSDFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 797
Qy      381 QAEKFPSEBNRLLLLHGFLENVHFHTSILLSFLVRACKPYDLQIYPOERSIRVPEEG 440
Db      798 QAEKFPSEBNRLLLLHGFLENVHFHTSILLSFLVRACKPYDLQIYPOERSIRVPEEG 857
Qy      441 EHYELHLHYLOENLGSRIAALKVI 465
Db      858 EHYELHLHYLOENLGSRIAALKVI 882

RESULT 10

```

ABU07720  
 ID ABU07720 standard; Protein: 882 AA.  
 AC ABU07720;  
 DT 19-MAY-2003 (first entry)  
 DE Human serine protease HIPHUM46.  
 XX  
 KM Human; enzyme; HIPHUM46; serine protease; gene therapy; osteoarthritis;  
 KM serine protease activity; modulation; dipeptidyl peptidase activity;  
 KM musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;  
 KM Alzheimer's disease; paraspinal nerve palsy; Huntington's disease;  
 KM amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;  
 KM irritable bowel syndrome; type 1 diabetes; faecal incontinence; anaemia;  
 KM haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;  
 KM colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;  
 KM multiple sclerosis.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Region Location/Qualifiers  
 FT 259..260  
 FT /note="Paired glutamates of the beta propeller domain"  
 FT 739  
 FT /label= Catalytic\_serine\_residue  
 FT 817  
 FT /label= Catalytic\_aspartate\_residue  
 FT 849  
 FT /label= Catalytic\_histidine\_residue  
 FT Active-site  
 PN GB2374869-A.  
 XX  
 PD 30-OCT-2002.  
 XX  
 PF 22-JAN-2002; 2002GB-0001404.  
 XX  
 PR 23-JAN-2001; 2001GB-0001760.  
 XX  
 PA (GLAXO ) GLAXO GROUP LTD.  
 XX  
 PI Edbrooke MR, Lewis AP;  
 DR WPI; 2003-150703/15.  
 DR N-PSDB; ABX12255.  
 XX  
 PT Identifying modulators of serine protease activity useful for treating  
 PT musculoskeletal diseases, by contacting cell expressing a novel serine  
 PT protease polypeptide with a compound and monitoring serine protease  
 PT activity -  
 PS Claim 10; Page 26-29; 38pp; English.  
 XX  
 CC The invention relates to a method of identifying a substance that  
 CC modulates serine protease activity, comprising contacting a cell such as  
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a  
 CC virus, expressing a serine protease polypeptide (HIPHUM 46), or its  
 CC variant having dipeptidyl peptidase activity, or a serine protease  
 CC isolated from the cell with a test substance and monitoring for serine  
 CC protease activity. The method is useful for identifying a substance that  
 CC modulates serine protease activity. A modulator of the serine protease that  
 CC is useful in the manufacture of a medicament for treatment or prophylaxis of  
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus  
 CC infection, Alzheimer's disease, paraspinal nerve palsy, myotonic  
 CC dystrophy, Huntington's disease, amyotrophic lateral sclerosis.  
 CC Additional disease that may be treated using modulators of the serine  
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung  
 CC disease, type 1 diabetes, faecal incontinence, haemorrhoids, proctitis,  
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,  
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple  
 CC sclerosis. The present sequence represents the amino acid sequence of the  
 CC human serine protease HIPHUM46.  
 XX

Sequence 882 AA;  
 Query Match 97.6%; Score 2422; DB 24; Length 882;  
 Best Local Similarity 82.3%; Pred. No. 1.6e-230;  
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;  
 QY 1 TGTANPKVTFKMSSEIMDAAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKAWMSI 60  
 DB 318 TGTANPKVTFKMSSEIMDAAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKAWMSI 377  
 QY 61 LDRSQTROIIVLISELPIPVDDVMEFORLIESVDSVTPLIYEETDWINIHDF 120  
 DB 378 LDRSQTROIIVLISELPIPVDDVMEFORLIESVDSVTPLIYEETDWINIHDF 437  
 QY 121 HVPQSHHEEIEFTFASCEKTRHLYKTSILKESKYRSSGGLPAPSPKCPKEIEA 180  
 DB 438 HVPQSHHEEIEFTFASCEKTRHLYKTSILKESKYRSSGGLPAPSPKCPKEIEA 497  
 QY 181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLVYVYVNGEVTRLTDRGYSH 240  
 DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLVYVYVNGEVTRLTDRGYSH 557  
 QY 241 SCCTISQHCDFPISKYKNQKNPCHVSLYKLSPPDDPTCKTEFWATILDSAGPLPDYTP 300  
 DB 558 SCCTISQHCDFPISKYKNQKNPCHVSLYKLSPPDDPTCKTEFWATILDSAGPLPDYTP 617  
 QY 301 EIRFSESTGTFTLYGMLYKPHLOPKKXPTVLFYIGSPQ----- 340  
 DB 618 EIRFSESTGTFTLYGMLYKPHLOPKKXPTVLFYIGSPQVLVNNRPGKVKYFRLNTLA 677  
 QY 341 ----- 340  
 DB 678 SLGYVVVVIDNRGSGRGLKFECAFYKMKQGLEIDDQVEGLQYLASRYDFIDDRVGIGH 727  
 QY 341 -----VAIAGAPVTLMTIYDPTGYTERYVNGHDPDQGYLGSVAM 380  
 DB 738 WSYGYLSLMAIMQSRDIFRVALAGAPVTLMTIYDPTGYTERYVNGHDPDQGYLGSVAM 797  
 QY 381 QAEKFPSEPRKLLLLGFLDENVHPAHTSILSLFLRAGKPYLQIYPOERHSIRVPSG 440  
 DB 798 QAEKFPSEPRKLLLLGFLDENVHPAHTSILSLFLRAGKPYLQIYPOERHSIRVPSG 857  
 QY 441 EHYELHLHYLQENLGSRIALVKVI 465  
 DB 858 EHYELHLHYLQENLGSRIALVKVI 882  
 RESULT 11  
 AAE14337  
 ID AAE14337 standard; Protein: 580 AA.  
 XX  
 AC AAE14337;  
 DT 07-MAR-2002 (first entry)  
 DE Human protease PRS-2 protein.  
 XX  
 KM Human; protease; PRS-2; tranquilliser; gene therapy; vaccine; allergy;  
 KM infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
 KM atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
 KM gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
 KM epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;  
 KM hypertension; neurological disorder; Parkinson's disease; drug screening;  
 KM cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;  
 KM diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;  
 KM autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
 KM developmental disorder; reproductive disorder; infertility; diarrhoea;  
 KM dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200183775-A2.

PD 08-NOV-2001.  
 XX 04-MAY-2001; 2001WO-US14651.  
 PF 04-MAY-2001; 2000US-202082P.  
 XX 11-MAY-2000; 2000US-203566P.  
 PR 17-MAY-2000; 2000US-205803P.  
 PR 25-MAY-2000; 2000US-207477P.  
 PR 01-JUN-2000; 2000US-209402P.  
 XX (INCYTE GENOMICS INC.  
 XX Deleage AM, Lal P, Hafalia A, Patterson C, Wallia NK, Kearney L;  
 PI Triponlay CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;  
 PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM;  
 PI Reddy R, Yue H, Tang YT;  
 DR WPI; 2002-034518/04.  
 DR N-PSDB; AAD23843.  
 XX Novel human proteases and polynucleotides encoding the proteases,  
 PT useful for treating, diagnosing or preventing cell proliferative,  
 PT cardiovascular, autoimmune/inflammatory, neurological and developmental  
 PT disorders -  
 PT  
 PS Claim 1; Page 120-121; 151pp; English.  
 XX The invention relates to human proteases (PRTS1-14) and its corresponding  
 CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for  
 CC the diagnosis, treatment and prevention of disorders associated with  
 CC increased or decreased expression of PRTS. Examples of such disorders  
 CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
 CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders  
 CC (AIDS, Addison's disease, allergy, anemia, asthma, atopic dermatitis,  
 CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
 CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
 CC viral, bacterial, fungal, parasitic, protozoal and helminthic  
 CC infections); cardiovascular disorders (myocardial infarction, ischaemic  
 CC heart disease and hypertension); neurological disorders (epilepsy,  
 CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,  
 CC Parkinson's disease, stroke, mental disorders including mood, anxiety  
 CC and seasonal affective disorder and prion diseases); gastrointestinal  
 CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);  
 CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,  
 CC alopecia, scabies, insect bites and urticaria); reproductive disorder  
 CC (infertility, disruption of estrous and menstrual cycle and  
 CC gynaecomastia); and developmental disorders (renal tubular acidosis,  
 CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
 CC PRTS DNA is also in useful is gene therapy. PRTS and its immunogenic  
 CC fragments are useful for screening libraries of compounds in several drug  
 CC screening assays. The present sequence is human protease PRTS-2 protein.  
 XX  
 SQ Sequence 580 AA;  
 Query Match 90.4%; Score 2244.5; DB 23; Length 580;  
 Best Local Similarity 84.7%; Pred. No. 3.3e-213;  
 Matches 436; Conservative 4; Mismatches 18; Indels 57; Gaps 4;  
 QY 1 TGTANPKYTFKMSSEIMDAEGRIIDVIDELIOPFFILEGVEYIARAWTPEGKXANSI 60  
 DB 73 TGTANPKYTFKMSSEIMDAEGRIIDVIDELIOPFFILEGVEYIARAWTPEGKXANSI 132  
 QY 61 LLDROGTRLOIYLISPELFIPEVDYMERORLIESVPSVTLIYEETDIIWINIHDF 120  
 DB 133 LLDROGTRLOIYLISPELFIPEVDYMERORLIESVPSVTLIYEETDIIWINIHDF 192  
 QY 121 HVFPOSHEEIEIFIFASECKTGRHLKYITSLKSKYRSSGGLPAPS-DEKCPKIEBI 179  
 DB 193 HVFPOSHEEIEIFIFASECKTGRHLKYITSLKSKYRSSGGLPAPVTMMITPMSL 252  
 QY 180 AITSEMEWLGHGNSIQVDERLRLVYFEGTQDSPLEHLVYVSVNPEEVTRLIDRGYS 239  
 DB 253 GTPSC--MCVTHIVEIQVDEVRRLVYFEGTQDSPLEHLVYVSVNPEEVTRLIDRGYS 309

QY 240 HSOCISQHCDFPISKYSNOKNPHCVSLYKSSPEDDPCTCKTEFMAITLDSAGLPDYTP 299  
 DB 310 HSOCISQHCDFPISKYSNOKNPHCVSLYKSSPEDDPCTCKTEFMAITLDSAGLPDYTP 369  
 QY 300 PEIFSESTTGTFTLYGMLYKPHDLPQKKYPTVLFIYGSPQ----- 340  
 DB 370 PEIFSESTTGTFTLYGMLYKPHDLPQKKYPTVLFIYGSPQVQVNNRFGVKYFRLNTL 429  
 QY 341 -----VAIAGPVTLMIFYDTGTYERTYKGDONE 370  
 DB 430 ASIGYVVVVVIDNRGSCRLKEGAFKYMVAIAGPVTLMIFYDTGTYERTYKGDONE 489  
 QY 371 OGYYIGSVAMQAEKPESEPNRLILHGFLENVHFANTSILSLVRAKPYDLOIYPOE 430  
 DB 490 OGYYIGSVAMQAEKPESEPNRLILHGFLENVHFANTSILSLVRAKPYDLOIYPOE 545  
 QY 431 RSHIRPESGEHYELHLHYLOENIGSRITAAKVI 465  
 DB 546 RSHIRPESGEHYELHLHYLOENIGSRITAAKVI 580  
 RESULT 12  
 ABG61594  
 ID ABG61594 standard; Protein; 690 AA.  
 AC ABG61594;  
 DT 12-AUG-2002 (first entry)  
 DE Human DPR-1 splice variant #1.  
 XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;  
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KW dyskinesia; reproductive disorder; inflammatory disorder;  
 KW metabolic disorder.  
 XX Homo sapiens.  
 OS WO200231134-A2.  
 PN 18-APR-2002.  
 PD 12-OCT-2001; 2001WO-US31874.  
 PF 12-OCT-2001; 2000US-240117P.  
 PR (FERR) FERRING BV.  
 PA Qi S, Akinsanya KO, Riviere PJ, Junien J;  
 PI WPI; 2002-444178/47.  
 DR N-PSDB; ABK83325.  
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
 PT viral infections, cancers, allergies, neurological disorders, or pain  
 PS Disclosure: Page 59-61; 113pp; English.  
 XX The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related  
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,



CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 690 AA;

Query Match 72.9%; Score 1809.5; DB 23; Length 690;  
 Best Local Similarity 94.0%; Pred. No. 6,2e-170;  
 Matches 342; Conservative 6; Mismatches 3; Indels 13; Gaps 3;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGTPEGKYAMS1 60  
 DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGTPEGKYAMS1 377  
 QY 61 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 120  
 DB 378 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 437  
 QY 121 HVPQSHHEEIEFIFASECTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEE1A 180  
 DB 438 HVPQSHHEEIEFIFASECTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEE1A 497  
 QY 181 ITSGEMEVLAGHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVSYNPGEVTRLTDRGYSH 240  
 DB 498 ITSGEMEVLAGHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVSYNPGEVTRLTDRGYSH 557  
 QY 241 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFMAITLDSAGLPDYTPP 300  
 DB 558 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFMAITLDSAGLPDYTPP 617  
 QY 301 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFYIG-----GPOVALIAGAPV----- 348  
 DB 618 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFYIGRLLLLGPG-SLGGSSMIQDTRN 676

QY 349 TLM1 352  
 : :  
 DB 677 VIMV 680

RESULT 13

ID ABG61600 standard; Protein; 658 AA.

AC ABG61600;

DT 12-AUG-2002 (first entry)

DE Human DPRP-1 splice variant #7.

XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;  
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KM dyskinesia; reproductive disorder; inflammatory disorder;  
 KM metabolic disorder.

OS Homo sapiens.

XX WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

XX 12-OCT-2000; 2000US-240117P.

PR (FERR ) FERRING BV.

PA Qi S, Akinsanya KO, Riviere PJ, Juntien J;

XX

DR MPI; 2002-44478/47.

XX N-PSDB; ABK03331.

XX

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding

PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and

PT viral infections, cancers, allergies, neurological disorders, or pain

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Disclosure: Page 70-72; 113pp; English.

CC The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related  
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypertension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 658 AA;

Query Match 72.8%; Score 1808; DB 23; Length 658;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-170;  
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGTPEGKYAMS1 60  
 DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGTPEGKYAMS1 377  
 QY 61 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 120  
 DB 378 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 437  
 QY 121 HVPQSHHEEIEFIFASECTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEE1A 180  
 DB 438 HVPQSHHEEIEFIFASECTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEE1A 497  
 QY 181 ITSGEMEVLAGHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVSYNPGEVTRLTDRGYSH 240  
 DB 498 ITSGEMEVLAGHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVSYNPGEVTRLTDRGYSH 557  
 QY 241 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFMAITLDSAGLPDYTPP 300  
 DB 558 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFMAITLDSAGLPDYTPP 617  
 QY 301 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFYIG 338  
 DB 618 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFYIG 655

RESULT 14

ID ABG61596 standard; Protein; 661 AA.

AC ABG61596;

DT 12-AUG-2002 (first entry)

DE Human DPRP-1 splice variant #3.

XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;  
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KM dyskinesia; reproductive disorder; inflammatory disorder;  
 KM metabolic disorder.

XX Homo sapiens.  
OS  
XX MO200231134-A2.  
XX  
XX 18-APR-2002.  
XX  
XX 12-OCT-2001, 2001MO-US31874.  
XX  
XX 12-OCT-2000, 2000US-240117P.  
XX  
XX (FERR ) FERRING BV.  
XX  
XX Qi S, Akinsanya KO, Riviere PJ, Junien J;  
XX WPI, 2002-444178/47.  
XX N-PSDB; ABR83327.  
XX  
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
PT viral infections, cancers, allergies, neurological disorders, or pain  
PT -  
XX  
XX Disclosure; Page 63-65; 113pp; English.  
XX  
XX The present invention relates to the isolation of novel human serine  
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related  
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
CC and nucleic acids encoding them are useful for treating infections  
CC such as fungal, bacterial, protozoan and viral infections, particularly  
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, osteoporosis, angina pectoris,  
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
CC schizophrenia), and dyskinesias. These may also be used in discovering  
CC therapeutic agents for the treatment of reproductive, inflammatory and  
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.  
XX  
XX Sequence 661 AA;  
SQ

Query Match 72.8%; Score 1808; DB 23; Length 661;  
Best Local Similarity 100.0%; Pred. No. 8, 1e-170;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGTANPKVTFKXSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 60  
DB 318 TGTANPKVTFKXSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 377  
QY 61 LLDROSTRQIYLISPELFIPEVDDVMERORLIESVPSVPLIYEETD1MINIHDF 120  
DB 378 LLDROSTRQIYLISPELFIPEVDDVMERORLIESVPSVPLIYEETD1MINIHDF 437  
QY 121 HVPQSHHEEIEIFIPASECKTGFRHLKYITSILKESKYRSSGGLPAPSDFCPIKEEIA 180  
DB 438 HVPQSHHEEIEIFIPASECKTGFRHLKYITSILKESKYRSSGGLPAPSDFCPIKEEIA 497  
QY 181 ITSGEWEVLGRHGSNIQVDEVRRLVFEGETKOSPLEHLLVYVSYPGEVTRLLTDGYSH 240  
DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVFEGETKOSPLEHLLVYVSYPGEVTRLLTDGYSH 557  
QY 241 SCCISOHCPFIISKYSNOKNPHCVSLYKLSPPEDDCTCKTKEFWATILDSAGLPDYTPP 300  
DB 558 SCCISOHCPFIISKYSNOKNPHCVSLYKLSPPEDDCTCKTKEFWATILDSAGLPDYTPP 617  
QY 301 EIFSESTGTFTLYGMLYKPHDLPQPKYPTVLFIYGG 338  
DB 618 EIFSESTGTFTLYGMLYKPHDLPQPKYPTVLFIYGG 655

RESULT 15  
ABG61601

ID ABG61601 standard; Protein; 613 AA.  
XX  
XX ABG61601;  
AC  
XX 12-AUG-2002 (first entry)  
DT  
XX Human DPRP-1 splice variant #8.  
DE  
XX  
XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP,  
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;  
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
KM dyskinesia; reproductive disorder; inflammatory disorder;  
KM metabolic disorder.  
XX  
XX Homo sapiens.  
OS  
XX MO200231134-A2.  
XX  
XX 18-APR-2002.  
XX  
XX 12-OCT-2001, 2001MO-US31874.  
XX  
XX 12-OCT-2000, 2000US-240117P.  
XX  
XX (FERR ) FERRING BV.  
XX  
XX Qi S, Akinsanya KO, Riviere PJ, Junien J;  
XX WPI, 2002-444178/47.  
XX N-PSDB; ABR83327.  
XX  
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and  
PT viral infections, cancers, allergies, neurological disorders, or pain  
PT -  
XX  
XX Disclosure; Page 73-75; 113pp; English.  
XX  
XX The present invention relates to the isolation of novel human serine  
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related  
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
CC and nucleic acids encoding them are useful for treating infections  
CC such as fungal, bacterial, protozoan and viral infections, particularly  
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, osteoporosis, angina pectoris,  
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
CC schizophrenia), and dyskinesias. These may also be used in discovering  
CC therapeutic agents for the treatment of reproductive, inflammatory and  
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.  
XX  
XX Sequence 613 AA;  
SQ  
Query Match 62.0%; Score 1540; DB 23; Length 613;  
Best Local Similarity 100.0%; Pred. No. 2, 8e-143;  
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGTANPKVTFKXSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 60  
DB 318 TGTANPKVTFKXSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 377  
QY 61 LLDROSTRQIYLISPELFIPEVDDVMERORLIESVPSVPLIYEETD1MINIHDF 120  
DB 378 LLDROSTRQIYLISPELFIPEVDDVMERORLIESVPSVPLIYEETD1MINIHDF 437  
QY 121 HVPQSHHEEIEIFIPASECKTGFRHLKYITSILKESKYRSSGGLPAPSDFCPIKEEIA 180  
DB 438 HVPQSHHEEIEIFIPASECKTGFRHLKYITSILKESKYRSSGGLPAPSDFCPIKEEIA 497

Qy	181	ITSGEMEVLAGRHSNIQVDEVRLVYFEGTKDSPL	EHHLVVSYNPGEVTRLTDRGYSH	240
Db	498	ITSGEMEVLAGRHSNIQVDEVRLVYFEGTKDSPL	EHHLVVSYNPGEVTRLTDRGYSH	557
Qy	241	SCCISOHCDFFISKYSNQKNPHCVSLYKLS	SPEDDPTCKTKEFWATILDS	230
Db	558	SCCISOHCDFFISKYSNQKNPHCVSLYKLS	SPEDDPTCKTKEFWATILDS	607

Search completed: October 15, 2003, 17:10:46  
Job time : 30.3565 secs



US-09-462-284-2  
 ; Sequence 2, Application US/09462284  
 ; Patent No. 6309868  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nestec S.A.  
 ; APPLICANT: Monod, Michel  
 ; APPLICANT: Doumas, Agnes  
 ; APPLICANT: Affolter, Michael  
 ; APPLICANT: Van den Broek, Peter  
 ; TITLE OF INVENTION: CLONING OF THE  
 ; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM  
 ; FILE REFERENCE: 8265-298  
 ; CURRENT APPLICATION NUMBER: US/09/462,284  
 ; CURRENT FILING DATE: 2000-01-03  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 771  
 ; TYPE: PRF  
 ; ORGANISM: Fungus  
 US-09-462-284-2

Query Match 12.5%; Score 309.5; DB 4; Length 771;  
 Best Local Similarity 23.3%; Pred. No. 2.1e-24;  
 Matches 128; Conservative 67; Mismatches 174; Indels 181; Gaps 20;

QY 5 NPKVTFKMSSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGWTEGKYAMSILDR 64  
 DB 259 NPTVLSS-----LNTASKEVKAQPIDAESTDLII-----GEVAM--LDT 298  
 QY 65 SOTRLQIVLISPELPIVEDDVMERORLIESVPSVTPLLIYEETDII--NINIDIFHV 123  
 DB 299 HTT-----VAAKAFNRVOD--QOKVAVDTASNKAT--VISDRDGTDLGDLMLSMKYT 349  
 QY 124 POSHEEIEIFIFASECKGFRLHYKITSLKESKXKRSSGGLPAPSDFCPIKEIATIS 183  
 DB 350 PKPSDKAYIYIDISDHSGMAHLVLP-----VSGGEPPI-----LTK 387  
 QY 184 GEMEVLGHRGNSNIQVDEVRLVYFEGTKDSPLEHLLVYVSVNPGEVRLTDRGYS--HS 241  
 DB 368 GDMET-----SILSIDQERQLVYVLTSTGHSTERNHLSYSY--STFAVPLVDVYAAKWS 442  
 QY 242 CCIISCHDFEISKYNSQKNPHCVSLYKLSPEDDPTCKTKEFWATILDSAGPL---PDYT 298  
 DB 443 ASFSANSGYVILTYGGPDVPR--QELVYTTWS-----TKPL--RITDNAKVLEQIKDYA 492  
 QY 299 PPEISFE--STTGFTLYGMLYKPHDLQPGKKYPTVLFYIGRP----- 339  
 DB 493 LPNITYFELPLPSGFTLVNMORLPFGFSPDKKYPILFTPYGGPGAQEVTKRMQALNPKAY 552  
 QY 340 ----- 339  
 DB 553 VASDESELEVYVTVNDRGTGFKGRKRAVTRQLGLLEAEDQIYAAQQAANIPIWDADHI 612  
 QY 340 -----OVAIAGAPVTLMIFYDTGTYTERYMGHPDNEGGYVYG 376  
 DB 613 GIMGWSFGGYLTSLKYLEKDSGAFITGVITAPYSMDRFDSMTYTERYMKTLSTNEEGYETS 672  
 QY 377 SVAMQAEKPPSEPNRLLLHGFLENVHPAHTSILSLFVRAKGKPYDQIYPOEHSIRV 436  
 DB 673 AV-RKTDGFKVVEGFLIQHGTGDDNVHFNQNSALVDLMDG-----VSFEKLHSQWF 725  
 QY 437 PESGEHYELH 446  
 DB 726 TDSHDGISYH 735

RESULT 3  
 US-10-002-593-6  
 ; Sequence 6, Application US/10002593  
 ; Patent No. 6586198  
 ; GENERAL INFORMATION:

; APPLICANT: Vanderbilt University  
 ; APPLICANT: Brown, Nancy J.  
 ; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTI  
 ; TITLE OF INVENTION: INHIBITOR AND VASOPRESSIN INHIBITOR ASSOCIATED ANGIOEDEMA  
 ; FILE REFERENCE: Acty Docket No. 6586198 1242/48/2  
 ; CURRENT APPLICATION NUMBER: US/10/002,593  
 ; CURRENT FILING DATE: 2001-10-31  
 ; PRIOR APPLICATION NUMBER: 60/244,524  
 ; PRIOR FILING DATE: 2000-10-31  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patencin version 3.1  
 ; SEQ ID NO 6  
 ; LENGTH: 766  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 US-10-002-593-6

Query Match 9.6%; Score 238; DB 4; Length 766;  
 Best Local Similarity 21.1%; Pred. No. 1.4e-16;  
 Matches 120; Conservative 71; Mismatches 193; Indels 186; Gaps 24;

QY 2 GANPKVTFKMSSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGWTEGKYAMSIL 61  
 DB 260 GAVNPTVAFV--VNTDLSSTVNTATSIQIAPASMLI-GDHYLCDVTMA----- 306  
 QY 62 LDRSOTRLQIVLISPELPIVEDDVMERORLIESVPSVTPLLIYEETDII--NINIDIFH 121  
 DB 307 ---TGERISLQWL-----RIQNV--SVMDICQDESSGMW--NC----- 339  
 QY 122 VFPQSHHEIEIFIFASECKGFRLHYKITSLKESKXKRSSGGLPAPSDFCPI 174  
 DB 340 LVAROHIMSTTGWGRFPEPFTLDGNSFYKIIS--NEEGYRHI-----CYFOID 390  
 QY 175 IKEEIAISGEMEVLGHRGNSNIQVDEVRLVYFEGT--KDSPLEHLLVYVSVNPGEVRL 233  
 DB 391 KKDCTFITKGTWEYIG-----IEALTSPLYIISNEYKGMFGKRNLYKIQLIDYKVTCL 445  
 QY 234 TDRGYSHSCCIS--OHCDPFIKYSNOKNPH--C-----VSLYKLSPEDDPTCKTKEFWA 285  
 DB 446 -----SCELNPERCQYVSFSKEMAKYQLRCSGPELPYTLHSSVNDGLRVLED--N 497  
 QY 286 TILDSAGLPPYTPPE---IFSESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGRP--- 339  
 DB 498 SALDKM--LQNVQWPSKKLDFIILNETKF--WYOMILRPH--PDKSKYPLLDVVAAGPSCQ 553  
 QY 340 ----- 339  
 DB 554 KADIVFRLNMTATYLASTENIIVASFGDGRSGYQGDKIIMHAINRRLGTFVEVDQIEAARQF 613  
 QY 340 -----OVAIAGAPVTLMIFYDTGTYTERYMG-- 364  
 DB 614 SKMGFVDRKRIAIWGSYGVYVTSVMVLSGSGVFCGIAVAVPVRMEWYDVSYTERYVGL 673  
 QY 365 -HPQONEGGYLGVSAMQAEKPPSEPNRLLLHGFLENVHPAHTSILSLFVRAKGKPYD 423  
 DB 674 PTPEDNLDHYNSYVMSAENF--KQVEYLLIHGTADDNVHFNQNSALVDLMDG----- 731  
 QY 424 LQIYPOEHSIRVPESEGEHYELHLYLOE 453  
 DB 732 AMWYTTDEDHGIIASSTAQHIIYTHMSHFIKQ 761

RESULT 4  
 PCT-US93-07923-3  
 ; Sequence 3, Application PC/TUS9307923  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Morimoto, Chikao  
 ; APPLICANT: Schloesman, Stuart F.  
 ; APPLICANT: Tanaka, Toshiaki  
 ; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Fish & Richardson

[illegible]

```

Db      659  YMGLEPTPEDNLDHYNSITVMSRAEKF--KQVEYLLIHGTADNVHFFQSAQISKALVDVG 7116
Oy      420  KPYDLQIYQERHSIRVPESEGEHVLHLPLOE 453
Db      717  VDFQAMWYTDDEHGIASTAHQHIYTHMSHFIKQ 750

RESULT 5
PCT-US93-07923-2
/ Sequence 2, Application PC/TUS9307923
/ GENERAL INFORMATION:
/ APPLICANT: Morimoto, Chikao
/ APPLICANT: Schlossman, Stuart F.
/ APPLICANT: Tanaka, Toshiaki
/ TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: Fish & Richardson
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ COMPUTER: IBM PS/2 Model 502 or 55SX
/ OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
/ SOFTWARE: WordPerfect (Version 5.0)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/07923
/ FILING DATE: 19930819
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/934,162
/ FILING DATE: 21-AUG-1992
/ APPLICATION NUMBER: 07/832,211
/ FILING DATE: 06-FEB-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Frazer, Janis K.
/ REGISTRATION NUMBER: 34,819
/ REFERENCE/DOCKET NUMBER: 00530/055002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 542-5070
/ TELEFAX: (617) 542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 759
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
PCT-US93-07923-2

Query Match      9.4%, Score 234, DB 5, Length 759;
Best Local Similarity 20.9%, Pred. No. 3,7e-16;
Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

Oy      2  GTANPKYTFKSEIMIDAEGRILIVIKELIQPELIFEGVEYIARAGWTPEGKAMGIL 61
Db      253  GAVNPYTKFFV--VNTSLSSTVMTATSIQIYAPASMLI-GHYLCDVTWA----- 299
Oy      62  LDNSQTRLQIVLSPELFIPVEDDWMERQRLIESVPSVTPLLIYEETTDIWINIHDFH 121
Db      300  ---TOERISISLOWL-----BRICWY--SYMDICDYDESSGRW-NC----- 332
Oy      122  VFPOSHEEIEFIPASCKTGFRH-----LYKITSILVSKYKRSSSGGLPAPSDFCP 174
Db      333  LVARQHLEMTGTGAVGFRPSEPHFTLIDGMSFYKIIIS--NEEGYRHT-----CYFQID 383
Oy      175  IKEEIIATSGEWEVLGRHGSNIQVDEVRLVYFEGT-KDSPLREHLTVVVSVMNGEYTRL 233
Db      384  KKQDCTFTTKGWIEVIG-----IEALTSDYLYYISNEYKMGMPGGGRNLVYKI-----QL 429

```

```

Oy      234 TDGYSHSCCIS-----OHKDFISXSNQKPH---C---VSLYKLSDEDDPTGCTK 281
Dd      430 SD--YTKVLTLSCELNPERCOYVSFSEAKYUYQRCSGROLRYTLHSSVNDKGRVL 487
Oy      282 EFMAITLDSAGLPIDYTPPE--ISFESTTGFTLYGMLYKPHLDQSKKYPTVLYFYCG 338
Dd      488 ED-NSALDKM--LQNVQMSKKLDLFIINETKF-WYOMILPRH-FDKSKKRYLLLDVYAG 542
Oy      339 P----- 339
Dd      543 PCSQKADTVRLMWTATYLASTENIIVASFDRGSGVGDQDKIMHINRLRGTFVEEDQIEA 602
Oy      340 -----QVAIAGAPVTLMLFTDTGTYTER 361
Dd      603 ARQFSKMGFVDNKRRIAMGWSYGYVTSMVLGSGGVFPCGIIAVALPVSWMYYSVYTER 662
Oy      362 YMG--HPDQNEGGYGLGSVAMQAEKEFPSEPNRLLLHGFLEBENVFAFTSILLSFLVRAG 419
Dd      663 YMGLPFPEBDMLDHYRNVSTVMSRAENF--KQVEYLLIHGTAADNVNFAQSSAQISKALVDVG 720
Oy      420 KPYDLQIYPOERHSIRVPESGEHYELHLHLHYOE 453
Dd      721 VDFQAMMYTDDEDHGIASSTAHQHTYTHMSGHFTKQ 754

```

RESULT 6  
 US-08-230-491A-3  
 Sequence 3, Application US/08230491A  
 Patent No. 5587299  
 GENERAL INFORMATION:  
 APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;  
 APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.  
 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR  
 TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES  
 TITLE OF INVENTION: THEREOF  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FELFE & LYNCH  
 STREET: 805 THIRD AVENUE  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: USA  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WORDPERFECT - ASC II  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/230,491A  
 FILING DATE: 20-APRIL-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hanson, No. 5587299man D.  
 REGISTRATION NUMBER: 30,946  
 REFERENCE/DOCKET NUMBER: LUD 330  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 688-9200  
 TELEFAX: (212) 838-3884  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 766 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

Query Match 9.4%; Score 234; DB 1; Length 766;  
 Best Local Similarity 20.9%; Pred. No. 3, 8e-16;  
 Matches 120; Conservative 70; Mismatches 199; Indels 194; Gaps 25;

Cy 2 GTAPKVTFFKSEIMIDAGRIIDVIDKELLOPFILTEGVEYIARAQMTBCKYAMSL 61  
 Db 260 GAVNETVFFV--VNTDSLSSVTNNTSIQITAPASMLI-GHHYLCDDVMA----- 306

```

QY 62 LBSQNRLOQVILSPFLFVDEVDWMEQROLIESPROSVPRLLIYEETDOWIMIHFI 121
Db 307 ---TQERISIQW-----RIQW-SWMDICDYDESSGM-NC----- 339
QY 122 VPDOSHEEELFIIFASECTGFRH-----LYKITSILKESKYKRSQGLPARBDECR 174
Db 340 LVARQHEIMSTGJWVGRFRSEPHFTLLDONSFYKII-S-NEGVRHI-----CYFOID 390
QY 175 IKEIJAITSGEWVLVGRHSGNIQVDEVRALUYFEET-KDSPLENHLVUVSVNPGVETRL 233
Db 391 KKQCTFTTGTATMEVIG-----LEALTSLYLLIISNEYGMFGKRNLYKI-----QL 436
QY 234 TDRGYSHSCCIS-----QHCDFFISKYGNOKNPH-----C-----VSLYKLSPEEDPTCKT 281
Db 437 SD-YTKVTCLSCELNPERCQYUYSVFSKAKYUOLRSGPGLRYTLTHSSVNDKGRVL 494
QY 282 EFPAATILDSAGRLPRDYTPRE--IISFESTGTFTLYGMLYKPHLOQCKKYPTVLFYCG 338
Db 495 ED-NSALDKM-LQNVQWPSKKLDEIILNETKF-WYQMI1LRPH-FDKSKKYRLLDVLVAG 549
QY 339 P----- 339
Db 550 PCSQKADTVFRLNMAVTLASTENIIVASFDKSGSGYCDKIMAHINRLGTFVEVDQIEA 609
QY 340 -----QVALACAPVTLMIPTDGTGTER 361
Db 610 AROFSKKGFLVNDKRIAIWMSYGVGYVTSMVLGSSGQVFKCIIAIVAPSRREYDYSVTER 669
QY 362 YMG--HPDQNEOGYULGSVAMQAEKFPSEPNRLNLHGFLENDVHFAHTSIIILSFVRAG 419
Db 670 YMLLPFPEDNLDHYNRNSTYMSRAEKF--KQVEYLLIHGTADQNVNHFGQSQNISKALVDVG 727
QY 420 KPYDLDIYPOERKSTRVPSGSEHYELHLNHYLOE 453
Db 728 VDEQAMWYTDDEHGIAASSTAHQIYTHMSHIKQ 761

```

RESULT 7  
 US-08-619-280A-3  
 : Sequence 3, Application US/08619280A  
 : Patent No. 5767242  
 : GENERAL INFORMATION:  
 : APPLICANT: Zimmermann, Rainer; Park, John E.;  
 : APPLICANT: Rettig, Wolfgang; Old, Lloyd J.  
 : TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN  
 : TITLE OF INVENTION: ALPHA, AND USES THEREOF  
 : NUMBER OF SEQUENCES: 10  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Felle & Lynch  
 : STREET: 805 Third Avenue  
 : CITY: New York City  
 : STATE: New York  
 : COUNTRY: USA  
 : ZIP: 10022  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
 : COMPUTER: IBM PS/2  
 : OPERATING SYSTEM: PC-DOS  
 : SOFTWARE: Wordperfect  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/619,280A  
 : FILING DATE: 18-MARCH-1996  
 : CLASSIFICATION: 433  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 08/230,491  
 : FILING DATE: 20-APRIL-1994  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Hanson, No. 5767242man D.  
 : REGISTRATION NUMBER: 30,946  
 : REFERENCE/DOCKET NUMBER: LUD 5330.1  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 766 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-619-280A-3

## Query Match

9.4%: Score 234; DB 1; Length 766;  
 Best Local Similarity 20.9%; Pred. No. 3.8e-16;

Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

QY 2 GTANPKTTFKMSIIMDAEGRILVIDKELIQPELIFEGVEYIARAGWTECKYASITL 61  
 DB 260 GAVNPYTKFV--VNTDSLSTVNATSIOITAPASMLI-GDHYLCVDTWA----- 306  
 QY 62 LDRSQTRLQIVLISPELIFVEDVMEORLIESVDSVTPLLIYEETTDIWINIHDI 121  
 DB 307 ---TOERISLQWL-----RRIQNY--SVMDICDYDESSGRW-NC----- 339  
 QY 122 VFPQSHHEEIEFIFASECKTGFRH-----LYKITSILKESYKXSSGGLPAPSDPKP 174  
 DB 340 LVARQHIEMSTGTGAVGRFRSEPHFTLDGNSFYKILS--NEEGYRHI-----CYFOID 390  
 QY 175 IKEEIAITSGEMVILGRHGSNIQVDEVRRLVFEET--KDSPLBHLIYVSVNPGEVTRL 233  
 DB 391 KKOCFTITKGTWEVIG-----IEALTSIDLYIISNEKMGKMPGGRNLTKI-----QL 436  
 QY 234 TDRGYSHSCCIS-----QHCDFFISKYSNQKNPH--C-----VSLYKLSSPEDDPTCKTK 281  
 DB 437 SD--YTKVTCISCELNPERCOYVSFSKEAKYQLRCSGGLPLVTLHSSVNDKGLRVL 494  
 QY 282 EFNATILDSAGPLPDYTPPE---IFSFEETGTFLYGMLYKPHLDQPKKYPTVLFYGG 338  
 DB 495 ED-NSALDKM--LQNVQMPKSLDFIILNETKF-WYQMLLPH-FDKSKKYPLLDDVYAG 549  
 QY 339 P-----RRIQNY--SVMDICDYDESSGRW-NC----- 339  
 DB 550 PCSQKADTVFRLNMTATLASTENIIVASPDGSGSYGQDKIMHAINRRLGFEVEDQIEA 609  
 QY 340 -----OVALAGAPVTLMIFYDTGYTER 361  
 DB 610 AROFSKMGFVNDKRIAIWGSYGYVTSMVLGSGGVFKGCIAPVPSRMEYDSDVYTER 669  
 QY 362 YMG--HPDQNEQGYLGSAVMAQAEKFPSEPRRLLLHGFLDENVHFAHTSILSFLVRAG 419  
 DB 670 YMGLPTEPDNLDRNSTVMSRAENF--KOVEYLLHGTADNVHFOQSAQISKALVDVG 727  
 QY 420 KPYDLQIYPOERHSIRVPESEGEHYELHLHYLOE 453  
 DB 728 VDFQAMWYTDDEHGIASSTAHHQIYTHMSHFIKQ 761

## RESULT 8

US-08-940-391-3  
 ; Sequence 3, Application US/08940391  
 ; Patent No. 5965373

## GENERAL INFORMATION:

APPLICANT: Zimmermann, Rainer, Park, John E.;  
 APPLICANT: Rettig, Wolfgang; Old, Lloyd J.  
 TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Felle & Lynch  
 STREET: 805 Third Avenue  
 CITY: New York City  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
 COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS  
 SOFTWARE: Mordperfect  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/940,391  
 FILING DATE: 01-OCT-1997  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/619,280  
 FILING DATE: 18-MARCH-1996  
 APPLICATION NUMBER: 08/230,491  
 FILING DATE: 20-APRIL-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hanson, No. 5965373man D.  
 REGISTRATION NUMBER: 30,946  
 REFERENCE/DOCKET NUMBER: LUD 5330.1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 688-9200  
 TELEFAX: (212) 838-3884  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 766 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-940-391-3

## Query Match

9.4%: Score 234; DB 2; Length 766;  
 Best Local Similarity 20.9%; Pred. No. 3.8e-16;

Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

QY 2 GTANPKTTFKMSIIMDAEGRILVIDKELIQPELIFEGVEYIARAGWTECKYASITL 61  
 DB 260 GAVNPYTKFV--VNTDSLSTVNATSIOITAPASMLI-GDHYLCVDTWA----- 306  
 QY 62 LDRSQTRLQIVLISPELIFVEDVMEORLIESVDSVTPLLIYEETTDIWINIHDI 121  
 DB 307 ---TOERISLQWL-----RRIQNY--SVMDICDYDESSGRW-NC----- 339  
 QY 122 VFPQSHHEEIEFIFASECKTGFRH-----LYKITSILKESYKXSSGGLPAPSDPKP 174  
 DB 340 LVARQHIEMSTGTGAVGRFRSEPHFTLDGNSFYKILS--NEEGYRHI-----CYFOID 390  
 QY 175 IKEEIAITSGEMVILGRHGSNIQVDEVRRLVFEET--KDSPLBHLIYVSVNPGEVTRL 233  
 DB 391 KKOCFTITKGTWEVIG-----IEALTSIDLYIISNEKMGKMPGGRNLTKI-----QL 436  
 QY 234 TDRGYSHSCCIS-----QHCDFFISKYSNQKNPH--C-----VSLYKLSSPEDDPTCKTK 281  
 DB 437 SD--YTKVTCISCELNPERCOYVSFSKEAKYQLRCSGGLPLVTLHSSVNDKGLRVL 494  
 QY 282 EFNATILDSAGPLPDYTPPE---IFSFEETGTFLYGMLYKPHLDQPKKYPTVLFYGG 338  
 DB 495 ED-NSALDKM--LQNVQMPKSLDFIILNETKF-WYQMLLPH-FDKSKKYPLLDDVYAG 549  
 QY 339 P-----RRIQNY--SVMDICDYDESSGRW-NC----- 339  
 DB 550 PCSQKADTVFRLNMTATLASTENIIVASPDGSGSYGQDKIMHAINRRLGFEVEDQIEA 609  
 QY 340 -----OVALAGAPVTLMIFYDTGYTER 361  
 DB 610 AROFSKMGFVNDKRIAIWGSYGYVTSMVLGSGGVFKGCIAPVPSRMEYDSDVYTER 669  
 QY 362 YMG--HPDQNEQGYLGSAVMAQAEKFPSEPRRLLLHGFLDENVHFAHTSILSFLVRAG 419  
 DB 670 YMGLPTEPDNLDRNSTVMSRAENF--KOVEYLLHGTADNVHFOQSAQISKALVDVG 727  
 QY 420 KPYDLQIYPOERHSIRVPESEGEHYELHLHYLOE 453  
 DB 728 VDFQAMWYTDDEHGIASSTAHHQIYTHMSHFIKQ 761

## RESULT 9

US-09-794-236-1  
 ; Sequence 1, Application US/09794236



Patent No. 6337069  
GENERAL INFORMATION:  
APPLICANT: Grouzmann, Eric  
APPLICANT: Lacroix, Jean-Silvain  
APPLICANT: Monod, Michel  
TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis  
FILE REFERENCE: 81985/276823  
CURRENT APPLICATION NUMBER: US/09/794,236  
CURRENT FILING DATE: 2001-02-28  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patentin version 3.0  
SEQ ID NO: 1  
LENGTH: 766  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-794-236-1

Query Match 9.4%; Score 234; DB 4; Length 766;  
Best Local Similarity 20.9%; Pred. No. 3.8e-16;  
Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

QY 2 GTANPKVTFKMSLIMDAEGRHIDVIDKLIQPEFLEGEVYIARAGWTEGKAMSIL 61  
DB 260 GAVNFTVAFV--VNTDLSSTYNTATSIQITAPASMLI-GDHYLCDVTMA----- 306  
QY 62 LDRSQTRLQIVLISBELFIPVEDDYMERORLIESVPSVTPLIYEETDIIWINDIFH 121  
DB 307 --TOERISLOWL-----RRIQNY--SYMDICDVEDSSGRW-NC----- 339  
QY 122 VFPOSHHEIEFIFASECKTGRRH-----LYKTSILSKSKYRSSGGLPABDPKCP 174  
DB 340 LVAQOHIEMTSTTGWGRFPEPHFTLDGNSFYKII--NEEGYKH-----CYFOID 390  
QY 175 IKEEIAITSGEMVLEGRHSNIQVDEVRLVFEET--KDSPLEHLLVYVYNPGEVTRL 233  
DB 391 KKDCTFITKGTMEVIG-----IEALTSPLYIISNEHYKMGGRLLYKI-----QL 436  
QY 234 TDRGYSHSCIS-----OHCDFIISKYSNQNPH--C-----VSLYKLSPEDDPTCKT 281  
DB 437 SD--YTKVTCISCEINPERCOYYSVFSKAKYQOLRCSPGLPLYTLHSSVNDKGLRVL 494  
QY 282 EFMAITLDSAGPLPDYTPRE--IFSFESTGFTLYGMLYKPHDQPKKYPVYFIYGG 338  
DB 495 ED-NSALDKM--LQNVQMPKSLDFTIINETKF-WYQMLRPH-FDKSKKYPPLLDVYAG 549  
QY 339 P----- 339  
DB 550 PCSQKADTVFRILNMTYLASTENIIVASFDRGSGYQGDKIMHAINRLGTFEVEDQIEA 609  
QY 340 -----QVAIAGAPVTLMIFYDTGYTER 361  
DB 610 ARQFSKMGFVNDKRIAIWGSYGYVTSMVLGSGGVFCGIAVAPVSRMEYDVSVYTER 669  
QY 362 YMG--HPONEGCGYGLGVAMQAQEKFPSEPNRLILLHGFLENHFAHTSILLSTLYAG 419  
DB 670 YMLGPTPEPDNDLHDYNSITVMSAENF--KQVEYLIHIGADNVHFOQSOISKALVDVG 727  
QY 420 KPYDLQIYPOERHSIRVPESEGEHYELHLYLOE 453  
DB 728 VDFQAMWYTDDEDHGIASTAHQHIYTHMSHFIKQ 761

RESULT 10  
US-08-230-491A-2  
Sequence 2, Application US/08230491A  
Patent No. 5587299  
GENERAL INFORMATION:  
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;  
APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR  
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN, AND USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:  
ADDRESSEE: FELFE & LYNCH  
STREET: 805 THIRD AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT - ASC II  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/230,491A  
FILING DATE: 20-APRIL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5587299man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 330  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 638-3884  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 760 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-230-491A-2

Query Match 8.6%; Score 214; DB 1; Length 760;  
Best Local Similarity 20.2%; Pred. No. 5.8e-14;  
Matches 87; Conservative 53; Mismatches 127; Indels 164; Gaps 13;

QY 140 KTGFRHLKYKTSILSKSKYRSSGGLPABDPKCPICEEIAITSGEMVLEGRHSNIQVD 199  
DB 371 KDGYKTHIHYI-----KOTVENAIQITSGKMEAI-----NI--- 400  
QY 200 EVRLVYFEETKDSPLHLLVYVYNPGEVTRLTDRGYSHS-CCISOH-----CDFEIS 253  
DB 404 -----FRTYQDLSLFYSSNEFEYPPGRNRIYRISISYPPSKKCVTKHAKERCOYTTA 453  
QY 254 KYSNQKNPHCVSLY-----KLSPEDDPTCKTEKFMATILDSAGPLD-----YTPPEIFS 304  
DB 454 SFSYAKYVYALVYCGPGIPISTLHDGRTDGE--IKILEENKELALAKNIQLPKKEIK 509  
QY 305 FEFTTGFTLYGMLYKPHDQPKKYPVYFIYGGP----- 339  
DB 510 KLEVDITLWYKMLLPQFPDRSKKYPULIQVGGPCSQSVRSVAVNMWISTYLSKKEGVI 569  
QY 340 ----- 339  
DB 570 ALVDGRTAFOGDKLLVAVYRKLGVEVEVDQITAVRKFIEMGFIDEKRIAIWGSYGYV 629  
QY 340 -----QVAIAGAPVTLMIFYDTGYTERYMGHP--DQNECGYGLGVAMQAQEK 385  
DB 630 SSLAIASGTGLFKCGIIVAPVSVWEYASVYTERFMGLPTKDDNLHFKSTVVARAEYF 689  
QY 386 PSEPNRLILLHGFLENHFAHTSILLSPYRACKPYDLQIYPOERHSIRVPESEGEHYELHLYLOE 442  
DB 690 RNVD--YLLIHGTADDNVHFOHNSAQIAKALVNAQVDFOAMYSIDQNHGL-----SGLSTNH 743  
QY 443 YEHLHLYLOE 453  
DB 744 LYTHMTHFLKQ 754

RESULT 11  
US-08-619-280A-2  
Sequence 2, Application US/08619280A  
Patent No. 5767242  
GENERAL INFORMATION:  
APPLICANT: Zimmermann, Rainer; Park, John E.;

```

/ APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
/ TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
/ NUMBER OF INVENTION: ALPHA, AND USES THEREOF
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Felle & Lynch
/ STREET: 805 Third Avenue
/ CITY: New York City
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10022
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
/ COMPUTER: IBM PS/2
/ OPERATING SYSTEM: PC-DOS
/ SOFTWARE: Wordperfect
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/619,280A
/ FILING DATE: 18-MARCH-1996
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/230,491
/ FILING DATE: 20-APRIL-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hanson, No. 5767242man D.
/ REGISTRATION NUMBER: 30,946
/ REFERENCE/DOCKET NUMBER: LUD 5330.1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 688-9200
/ TELEFAX: (212) 838-3884
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 760 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/
/ US-08-619-280A-2
/
/ Query Match
/ Best Local Similarity 20.2%; Pred. No. 5.8e-14;
/ Matches 87; Conservative 53; Mismatches 127; Indels 164; Gaps 13;
/
/ QY 140 KTGPHLLKITSILKESKTKRSSGGLPAPSDPKCPKEIAITSGEWELGHRGSIQVD 199
/ Db 371 KDGYKHIIYI-----KDTVENAIOITSGKWEAI-----NI----- 400
/
/ QY 200 EVRRLVPEGTGKDSPLBHLVYVYNPGEVTRLDRGSHS-CCISQH-----CDFIS 253
/ Db 401 -----FRVTDLSLFYSSNEFEETPGRRNIRYISIGYPPSKKCVTLKREKCOQYTA 453
/
/ QY 254 KYSNQKNHCVSLY-----KLSSPEDDPTCKTEFWATILDSAGPLPD-----YTPPEIFS 304
/ Db 454 SFSDAKYVALVCGPGIPISLHDGRDQE-----IKILENKELLENALKNIOQLPEIK 509
/
/ QY 305 FESITGFLYGLMLYRPHLOPGKKYPTVLFIYGGP----- 339
/ Db 510 KLEVEITLWMYKMLPPOFDRSKKYPLIIOYGGPSSOSVRSVFAVNMISYLASKGMVI 569
/
/ QY 340 ----- 339
/ Db 570 ALVDRGTAFOGDKLLVAVYRKLGYEVEDQITAVRKFIEMGFIDEKRIAIWGSYGGYV 629
/
/ QY 340 -----QVAIAGAPVTLMIFYDTGYTERYMGHP--DQNEOGYILGSVAMQAEKF 385
/ Db 630 SSLALASGTGLFKGCIATAVAVPSSWEYASVYTERFMGLPTKODNLEHYKNSTVMARAEYF 689
/
/ QY 386 PSEPRKLLHGFLENDVHFAHTSILSLFLVPAKRPYDIQIYPOERHSIRVEG--EH 442
/ Db 630 RNVD--YLLIHGTADNVHFOHSAQIAKALVNAOVDFOAMWYSQDNHGL-----SGISTNH 743
/
/ QY 443 YELHLHYLOE 453
/ Db 744 LYTHMTHFLKO 754

```

```

/ RESULT 12
/ US-08-940-391-2
/ Sequence 2, Application US/08940391
/ Patent No. 5965373
/ GENERAL INFORMATION:
/ APPLICANT: Zimmermann, Rainer; Park, John E.;
/ TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
/ PROTEIN ALPHA, AND USES THEREOF
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Felle & Lynch
/ STREET: 805 Third Avenue
/ CITY: New York City
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10022
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
/ COMPUTER: IBM PS/2
/ OPERATING SYSTEM: PC-DOS
/ SOFTWARE: Wordperfect
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/940,391
/ FILING DATE: 01-OCT-1997
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/619,280
/ FILING DATE: 18-MARCH-1996
/ APPLICATION NUMBER: 08/230,491
/ FILING DATE: 20-APRIL-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hanson, No. 5965373man D.
/ REGISTRATION NUMBER: 30,946
/ REFERENCE/DOCKET NUMBER: LUD 5330.1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 688-9200
/ TELEFAX: (212) 838-3884
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 760 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/
/ US-08-940-391-2
/
/ Query Match
/ Best Local Similarity 20.2%; Pred. No. 5.8e-14;
/ Matches 87; Conservative 53; Mismatches 127; Indels 164; Gaps 13;
/
/ QY 140 KTGPHLLKITSILKESKTKRSSGGLPAPSDPKCPKEIAITSGEWELGHRGSIQVD 199
/ Db 371 KDGYKHIIYI-----KDTVENAIOITSGKWEAI-----NI----- 400
/
/ QY 200 EVRRLVPEGTGKDSPLBHLVYVYNPGEVTRLDRGSHS-CCISQH-----CDFIS 253
/ Db 401 -----FRVTDLSLFYSSNEFEETPGRRNIRYISIGYPPSKKCVTLKREKCOQYTA 453
/
/ QY 254 KYSNQKNHCVSLY-----KLSSPEDDPTCKTEFWATILDSAGPLPD-----YTPPEIFS 304
/ Db 454 SFSDAKYVALVCGPGIPISLHDGRDQE-----IKILENKELLENALKNIOQLPEIK 509
/
/ QY 305 FESITGFLYGLMLYRPHLOPGKKYPTVLFIYGGP----- 339
/ Db 510 KLEVEITLWMYKMLPPOFDRSKKYPLIIOYGGPSSOSVRSVFAVNMISYLASKGMVI 569
/
/ QY 340 ----- 339
/ Db 570 ALVDRGTAFOGDKLLVAVYRKLGYEVEDQITAVRKFIEMGFIDEKRIAIWGSYGGYV 629
/
/ QY 340 -----QVAIAGAPVTLMIFYDTGYTERYMGHP--DQNEOGYILGSVAMQAEKF 385
/ Db 630 SSLALASGTGLFKGCIATAVAVPSSWEYASVYTERFMGLPTKODNLEHYKNSTVMARAEYF 689

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:13:57 : Search time 65.704 Seconds  
(without alignments)  
1140.337 Million cell updates/sec

Title: US-10-070-464-5

Perfect score: 2482

Sequence: 1 TGTANPKVTFPKSEIMIDAE.....HLHLYGENTGSRIALKVI 465

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2422	97.6	882	10	US-09-976-674-1 Sequence 1, Appli
2	2422	97.6	882	12	US-10-054-776-2 Sequence 2, Appli
3	2422	97.6	882	12	US-10-170-789-38 Sequence 38, Appli
4	1809.5	72.9	690	10	US-09-976-674-7 Sequence 7, Appli
5	1808	72.8	658	10	US-09-976-674-19 Sequence 19, Appli
6	1808	72.8	661	10	US-09-976-674-11 Sequence 11, Appli
7	1540	62.0	613	10	US-09-976-674-21 Sequence 21, Appli
8	1494.5	60.2	863	10	US-09-976-674-3 Sequence 3, Appli
9	1494.5	60.2	892	10	US-09-976-674-23 Sequence 23, Appli
10	1494.5	60.2	892	10	US-09-976-674-27 Sequence 27, Appli
11	1445	58.2	879	10	US-09-976-674-33 Sequence 33, Appli
12	1445	58.2	879	10	US-09-976-674-35 Sequence 35, Appli
13	1377.5	55.5	510	12	US-09-993-959-4 Sequence 4, Appli
14	1187.5	47.8	508	15	US-10-237-271-3 Sequence 3, Appli
15	1094.5	44.1	832	10	US-09-976-674-29 Sequence 29, Appli

16	1094.5	44.1	832	10	US-09-976-674-31 Sequence 31, Appli
17	1045	42.1	819	10	US-09-976-674-37 Sequence 37, Appli
18	1045	42.1	819	10	US-09-976-674-39 Sequence 39, Appli
19	471.5	19.0	518	10	US-09-976-674-25 Sequence 25, Appli
20	321	12.9	129	11	US-09-764-891-3564 Sequence 3564, Ap
21	277	11.2	818	12	US-10-401-437-3 Sequence 3, Appli
22	277	11.2	818	12	US-10-402-312-3 Sequence 3, Appli
23	277	11.2	818	12	US-10-402-067-3 Sequence 3, Appli
24	277	11.2	818	12	US-10-401-436-3 Sequence 3, Appli
25	267	10.8	93	9	US-09-867-550-988 Sequence 988, App
26	249.5	10.1	767	15	US-10-165-603-4 Sequence 4, Appli
27	247	10.0	706	10	US-09-976-674-41 Sequence 41, Appli
28	247	10.0	796	10	US-09-976-674-5 Sequence 5, Appli
29	247	10.0	796	12	US-09-870-133-2 Sequence 2, Appli
30	247	10.0	796	15	US-10-160-501-5 Sequence 5, Appli
31	238	9.6	766	12	US-10-423-714-6 Sequence 6, Appli
32	238	9.6	766	14	US-10-002-593-6 Sequence 6, Appli
33	238	9.6	766	15	US-10-165-603-7 Sequence 7, Appli
34	234	9.4	766	9	US-09-265-606-3 Sequence 3, Appli
35	234	9.4	766	12	US-09-993-959-1 Sequence 1, Appli
36	214	8.6	760	9	US-09-265-606-2 Sequence 2, Appli
37	214	8.6	760	12	US-10-301-822-55 Sequence 55, Appli
38	214	8.6	760	15	US-10-177-293-136 Sequence 136, App
39	214	8.6	803	12	US-10-205-219-58 Sequence 58, Appli
40	211.5	8.5	691	10	US-09-976-674-43 Sequence 43, Appli
41	190.5	7.7	710	15	US-10-156-761-10681 Sequence 10681, A
42	165	6.6	164	15	US-10-106-698-6740 Sequence 6740, Ap
43	146	5.9	432	15	US-10-237-271-4 Sequence 4, Appli
44	112	4.5	358	10	US-09-976-674-13 Sequence 13, Appli
45	104	4.2	1436	9	US-09-815-242-5566 Sequence 5566, Ap

#### ALIGNMENTS

RESULT 1	
US-09-976-674-1	
Sequence 1, Application US/09976674	
Patent No. US20020115843A1	
GENERAL INFORMATION:	
APPLICANT: Qi, Steve	
APPLICANT: Akinsanya, Karen	
APPLICANT: Riviere, Pierre	
APPLICANT: Junien, Jean-Louis	
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV	
FILE REFERENCE: 70669	
CURRENT APPLICATION NUMBER: US/09/976,674	
CURRENT FILING DATE: 2001-10-12	
PRIOR APPLICATION NUMBER: US 60/240,117	
PRIOR FILING DATE: 2000-10-12	
NUMBER OF SEQ ID NOS: 61	
SOFTWARE: PatentIn version 3.1	
SEQ ID NO 1	
LENGTH: 882	
TYPE: PRT	
ORGANISM: Homo sapiens	
US-09-976-674-1	
Query Match	97.6% Score 2422; DB 10; Length 882;
Best Local Similarity	82.3% Pred. No. 1.4e-231;
Matches	465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;
QY	1 TGTANPKVTFPKSEIMIDAEGRRIIDVDKELIOPELIFEGVEYIARAGMTPEGKYAWSI 60
DB	318 TGTANPKVTFPKSEIMIDAEGRRIIDVDKELIOPELIFEGVEYIARAGMTPEGKYAWSI 377
QY	61 LLDSSQRLQVLSPFLFVEDDWMERGLISSVPSVTPLIYEETIWINIHIF 120
DB	378 LLDSSQRLQVLSPFLFVEDDWMERGLISSVPSVTPLIYEETIWINIHIF 437
QY	121 HVPQSHHEEIEFTFASCKTGFRHLKYTSLKESKYKRSAGLPAAPSDPCPIKEBIA 180
DB	438 HVPQSHHEEIEFTFASCKTGFRHLKYTSLKESKYKRSAGLPAAPSDPCPIKEBIA 497

QY 181 ITSGEWELVGRHGSNIQVDEVRLLVFEETKDSPLEHNLVYVSVNPGSVTRLDGYSYH 240  
| | | | |  
DB 498 ITSGEWELVGRHGSNIQVDEVRLLVFEETKDSPLEHNLVYVSVNPGSVTRLDGYSYH 557  
QY 241 SCCISQHCDFPISKYSNQNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLDPTYP 300  
| | | | |  
DB 556 SCCISQHCDFPISKYSNQNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLDPTYP 617  
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKYPVLYFYGGPQVLVNNRFKGVYFRILNTLA 340  
| | | | |  
DB 618 EIFSFESTTGFTLYGMLYKPHDLQPGKYPVLYFYGGPQVLVNNRFKGVYFRILNTLA 677  
QY 341 -----VAIAGAPVTLMTFYDTGYTERYMGHPDQNEOGYIYLGVSVM 380  
| | | | |  
DB 678 SLGYVVVVIDNRGSGHRLKFEAGFYKMGQIEIDDOVEGLQYLASRYDFIDLDVYING 737  
QY 341 -----VAIAGAPVTLMTFYDTGYTERYMGHPDQNEOGYIYLGVSVM 380  
| | | | |  
DB 738 WSYGYLSLMAQMSDIFRVALIAGAPVTLMTFYDTGYTERYMGHPDQNEOGYIYLGVSVM 797  
QY 381 QAEKPESEPNRLLHLHGFLENVFAHTSILSLFLVRACKPYDLOIYPOERHSIRVPESG 440  
| | | | |  
DB 798 QAEKPESEPNRLLHLHGFLENVFAHTSILSLFLVRACKPYDLOIYPOERHSIRVPESG 857  
QY 441 EHYELHLHYLOENLGSRIALAKVI 465  
| | | | |  
DB 858 EHYELHLHYLOENLGSRIALAKVI 882

RESULT 2  
US-10-054-776-2

Sequence 2, Application US/10054776  
Publication No. US20030165818A1  
GENERAL INFORMATION:  
APPLICANT: Mark Robert Edbrooke  
APPLICANT: Alan Peter Lewis  
TITLE OF INVENTION: NOVEL PROTEIN  
FILE REFERENCE: OG1042US  
CURRENT APPLICATION NUMBER: US/10/054, 776  
CURRENT FILING DATE: 2002-01-23  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 882  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-054-776-2

Query Match 97.6%; Score 2422; DB 12; Length 882;  
Best Local Similarity 82.3%; Pred. No. 1.4e-231;  
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;  
QY 1 TGTANPKYTFKMSIMIDAGRIIDVDELIOPFELPEGVEYIARAGTPEGKAWSI 60  
| | | | |  
DB 318 TGTANPKYTFKMSIMIDAGRIIDVDELIOPFELPEGVEYIARAGTPEGKAWSI 377  
QY 61 LLDRSQTRLIQVILISPELFIPVEDDVMEORLIESVPDSVTPLLIYEETDIIWINDHF 120  
| | | | |  
DB 378 LLDRSQTRLIQVILISPELFIPVEDDVMEORLIESVPDSVTPLLIYEETDIIWINDHF 437  
QY 121 HVPFQSHHEEIEIFIPASECTGFRHLKYTSTILKESKYRSGGLPAPDFKCPIKEEIA 180  
| | | | |  
DB 438 HVPFQSHHEEIEIFIPASECTGFRHLKYTSTILKESKYRSGGLPAPDFKCPIKEEIA 497  
QY 181 ITSGEWELVGRHGSNIQVDEVRLLVFEETKDSPLEHNLVYVSVNPGSVTRLDGYSYH 240  
| | | | |  
DB 498 ITSGEWELVGRHGSNIQVDEVRLLVFEETKDSPLEHNLVYVSVNPGSVTRLDGYSYH 557  
QY 241 SCCISQHCDFPISKYSNQNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLDPTYP 300  
| | | | |  
DB 556 SCCISQHCDFPISKYSNQNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLDPTYP 617

QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKYPVLYFYGGPQVLVNNRFKGVYFRILNTLA 340  
| | | | |  
DB 618 EIFSFESTTGFTLYGMLYKPHDLQPGKYPVLYFYGGPQVLVNNRFKGVYFRILNTLA 677  
QY 341 -----VAIAGAPVTLMTFYDTGYTERYMGHPDQNEOGYIYLGVSVM 380  
| | | | |  
DB 678 SLGYVVVVIDNRGSGHRLKFEAGFYKMGQIEIDDOVEGLQYLASRYDFIDLDVYING 737  
QY 341 -----VAIAGAPVTLMTFYDTGYTERYMGHPDQNEOGYIYLGVSVM 380  
| | | | |  
DB 738 WSYGYLSLMAQMSDIFRVALIAGAPVTLMTFYDTGYTERYMGHPDQNEOGYIYLGVSVM 797  
QY 381 QAEKPESEPNRLLHLHGFLENVFAHTSILSLFLVRACKPYDLOIYPOERHSIRVPESG 440  
| | | | |  
DB 798 QAEKPESEPNRLLHLHGFLENVFAHTSILSLFLVRACKPYDLOIYPOERHSIRVPESG 857  
QY 441 EHYELHLHYLOENLGSRIALAKVI 465  
| | | | |  
DB 858 EHYELHLHYLOENLGSRIALAKVI 882

RESULT 3  
US-10-170-789-38

Sequence 38, Application US/10170789  
Publication No. US20030180930A1  
GENERAL INFORMATION:

APPLICANT: Rachel E. Meyers  
APPLICANT: Olandt, Peter J.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
AND PROTEASE FAMILY MEMBERS AND USES THEREOF  
FILE REFERENCE: 10448-191001  
CURRENT APPLICATION NUMBER: US/10/170, 789  
CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797, 039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186, 061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882, 166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212, 078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934, 406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226, 740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861, 801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205, 508  
PRIOR FILING DATE: 2000-05-19  
PRIOR APPLICATION NUMBER: US 09/801, 267  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07138  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187, 454  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/829, 671  
PRIOR FILING DATE: 2001-04-10  
PRIOR APPLICATION NUMBER: PCT/US01/40483  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: US 60/197, 508  
PRIOR FILING DATE: 2000-04-18

```

; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-789-38
```

```

Query Match      97.6%; Score 2422; DB 12; Length 882;
Best Local Similarity 82.3%; Pred. No. 1,4e-231;
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;
```

```

QY 1 TGTANPKYTFKMSKSEIMDAEGRIIDVIDKELIQPEILLFEGVEYIARAGWTEGKYANSI 60
DB 318 TGTANPKYTFKMSKSEIMDAEGRIIDVIDKELIQPEILLFEGVEYIARAGWTEGKYANSI 377
QY 61 LLDRSQTRLQIVLISPELFIPEDDVMERORLIESVPSVPLIYEETDIDMINIHDF 120
DB 378 LLDRSQTRLQIVLISPELFIPEDDVMERORLIESVPSVPLIYEETDIDMINIHDF 437
QY 121 HVPFQSHHEEIEFIASCKTGFRHLKYITSLKSKYKSSGGLPAPSDFCPIKEIA 180
DB 438 HVPFQSHHEEIEFIASCKTGFRHLKYITSLKSKYKSSGGLPAPSDFCPIKEIA 497
QY 181 ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVSYNPGEVRLTDRGYSH 240
DB 498 ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVSYNPGEVRLTDRGYSH 557
QY 241 SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 300
DB 558 SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 617
QY 301 EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQ-----GFOVAIAGAPV----- 340
DB 618 EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQVUNNRKGVKVFRLNTLA 677
QY 341 ----- 340
DB 678 SLGYVVVVINDRGSCHRLKEGAFKYMKGQIEIDVQGLQYLASRYDFILDREVIGH 737
QY 341 -----VAIAGAPVTLWTFYVYTGTERYMGHPDQNEGYIYGSYAM 380
DB 738 WSYGYISLWALMORSDIFRVAIAGAPVTLWTFYDTGTERYMGHPDQNEGYIYGSYAM 797
QY 381 QAEKFPSEPNRLLLHGLDENVFAHTSILSLFVRAGKPYDQIQOERHSIRVPSG 440
DB 798 QAEKFPSEPNRLLLHGLDENVFAHTSILSLFVRAGKPYDQIQOERHSIRVPSG 857
QY 441 EHYELHLHYIQENUGSRIALKVI 465
DB 858 EHYELHLHYIQENUGSRIALKVI 882
```

```

RESULT 4
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-7
```

```

Query Match      72.9%; Score 1809.5; DB 10; Length 690;
Best Local Similarity 94.0%; Pred. No. 8.4e-171;
Matches 342; Conservative 6; Mismatches 3; Indels 13; Gaps 3;
```

```

QY 1 TGTANPKYTFKMSKSEIMDAEGRIIDVIDKELIQPEILLFEGVEYIARAGWTEGKYANSI 60
DB 318 TGTANPKYTFKMSKSEIMDAEGRIIDVIDKELIQPEILLFEGVEYIARAGWTEGKYANSI 377
QY 61 LLDRSQTRLQIVLISPELFIPEDDVMERORLIESVPSVPLIYEETDIDMINIHDF 120
DB 378 LLDRSQTRLQIVLISPELFIPEDDVMERORLIESVPSVPLIYEETDIDMINIHDF 437
QY 121 HVPFQSHHEEIEFIASCKTGFRHLKYITSLKSKYKSSGGLPAPSDFCPIKEIA 180
DB 438 HVPFQSHHEEIEFIASCKTGFRHLKYITSLKSKYKSSGGLPAPSDFCPIKEIA 497
QY 181 ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVSYNPGEVRLTDRGYSH 240
DB 498 ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVSYNPGEVRLTDRGYSH 557
QY 241 SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 300
DB 558 SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 617
QY 301 EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQ-----GFOVAIAGAPV----- 348
DB 618 EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQLLLGPO-SLCCSSMIODTRN 676
QY 349 TLWI 352
DB 677 VIWV 680
```

```

RESULT 5
US-09-976-674-19
; Sequence 19, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-976-674-19

Query Match 72.8%; Score 1808; DB 10; Length 658;

Best Local Similarity 100.0%; Pred. No. 1,1e-170; Indels 0; Gaps 0;

Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGTANPKVTFKMSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 60
DB 318 TGTANPKVTFKMSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 377
QY 61 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 120
DB 378 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 437
QY 121 HVPFOSHEEIEIFIFASECKTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEEIA 180
DB 438 HVPFOSHEEIEIFIFASECKTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEEIA 497
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDGYSH 240
DB 498 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDGYSH 557
QY 241 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFWMATILDSAGPLPDYTPP 300
DB 558 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFWMATILDSAGPLPDYTPP 617
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 338
DB 618 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 655
```

RESULT 6

US-09-976-674-11

Sequence 11, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akimsanya, Karen

APPLICANT: Riviere, Pierre

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.1

SEQ ID NO 11

LENGTH: 661

TYPE: PRT

ORGANISM: Homo sapiens

US-09-976-674-11

Query Match 72.8%; Score 1808; DB 10; Length 661;

Best Local Similarity 100.0%; Pred. No. 1,1e-170; Indels 0; Gaps 0;

Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGTANPKVTFKMSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 60
DB 318 TGTANPKVTFKMSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 377
QY 61 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 120
DB 378 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 437
QY 121 HVPFOSHEEIEIFIFASECKTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEEIA 180
DB 438 HVPFOSHEEIEIFIFASECKTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEEIA 497
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDGYSH 240
DB 498 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDGYSH 557
```

```
QY 241 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFWMATILDSAGPLPDYTPP 300
DB 558 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFWMATILDSAGPLPDYTPP 617
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 338
DB 618 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 655
```

RESULT 7

US-09-976-674-21

Sequence 21, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akimsanya, Karen

APPLICANT: Riviere, Pierre

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.1

SEQ ID NO 21

LENGTH: 613

TYPE: PRT

ORGANISM: Homo sapiens

US-09-976-674-21

Query Match 62.0%; Score 1540; DB 10; Length 613;

Best Local Similarity 100.0%; Pred. No. 4,6e-144; Indels 0; Gaps 0;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGTANPKVTFKMSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 60
DB 318 TGTANPKVTFKMSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 377
QY 61 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 120
DB 378 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 437
QY 121 HVPFOSHEEIEIFIFASECKTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEEIA 180
DB 438 HVPFOSHEEIEIFIFASECKTGFRHLKYKTSILKESKYRSSGGLPAPSDFCPIKEEIA 497
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDGYSH 240
DB 498 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDGYSH 557
QY 241 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFWMATILDS 290
DB 558 SCCISQHCDFPISKYSNQNPKHCVSLYKLSPEDDPTCKTKFWMATILDS 607
```

RESULT 8

US-09-976-674-3

Sequence 3, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akimsanya, Karen

APPLICANT: Riviere, Pierre

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12



```

; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 863
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-976-674-3

```

Query Match	60.2%;	Score 1494.5;	DB 10;	Length 863;
Best Local Similarity	50.5%;	Pred. No. 2.6e-139;		
Matches 281;	Conservative 69;	Mismatches 105;	Indels 101;	Gaps 2

```

QY      TGANKPVFKXSEIMIDAEGRIDVIDEKLLOPEILLPEGEVYIARAGGTEPKXVMSI 60
Db      308 TGSKNPFIALKIAEFOTDQSKIYSTOEKELVQPPSLPKVEYIARAGGTRDKYAMAM 367
QY      61 LDRSQTRLOIYLISBELFIPEVDYMERKORLIESVPDVPRLIYEETTDIMINIDIF 120
Db      368 FLDRPQWLOIQLVLLRPALFIPSTENEGOLAGARAVPRNVQRYVVEEVTVMWINVDIF 427
QY      121 HVYPOSH-EEELFIYASCEKTGFKNLYKITSILKSKYKSSGGLPAPSDFKCPRIEEL 179
Db      428 YRPPOSEGBDELCEFLANCKCTGFCCHLYKVTVLVSOGYDWSSEPPSPGDEFFKCPRIEEI 487
QY      180 AITSGMEVLGRHGSNIQVDEVRRLVYEGCTDSPLLENHLYVVSVANPGEVYRLTDGXS 239
Db      488 ALTSGMEVLARHGSKIWNNEETKLVPFGTADTLEHNLVVSIEAAGEIVRLTTGFS 547
QY      240 HSCCISOHCDFFISKSNOKNHCUSLYLSSPEDDPTCKTKEPMATILDSAGLPDYTP 299
Db      548 HGSMSQNDMFVSHNSYSTPRCHNVYLLSPRDDRLHKQPRWASMMEAASCPDYVP 607
QY      300 PEIFSESTTGFTLYGMLYKPHDLOPKKKYPTVLFIYSGPO----- 340
Db      608 PEIFEHTRSVDRLYGMIKYPRHALQPKKHPVLVFGVQVQVLVNNSPKGIKYLRLNTL 667
QY      341 ----- 340
Db      668 ASLGVAVVVIDGRSGCQRLREGALKIOMQGOVELEDQLOFAVEKYGFTDLRSVAIH 727
QY      341 -----VAIAGAPVTLWIFPDYGTYSRWYGNHPDQNOEGYULASVA 379
Db      728 GMSYQVLIAMGLIHNGIHPQVFKYALAGAPVTVMAADTYTTRVYDVENBNQHTEASVA 787
QY      380 MQAEKFPSEPNRLLLLHGFEDENVAHFAHTSILSFLVAGKRYDLOIYPOERHSIRVES 439
Db      788 LHVEKLPENPNRLLIHGFLEDENVAHFPHNFLVSQLIAGKRPYDLOIYPERHSIRCPES 847
QY      440 GEHYELHLHYOENL 455
Db      848 GEHYEVTLLHAFLOEYLI 863

```

```

RESULT 9
US-09-976-674-23
; Sequence 23, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Ol, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 892
; TYPE: PRT

```

ORGANISM: Homo sapiens  
US-09-976-674-23

Query Match	60.2%	Score 1494.5;	DB 10;	Length 892;
Best Local Similarity	50.5%;	Pred. NO. 2.7e-139;		
Matches 281; Conservative	69;	Mismatches 105;	Indels 101;	Gaps 2

Qy	1	TGRANKVTPFKMSEIMDMBEGRIIDVIDKELIOPELLFEGVEVYIAAAGWTPBECKYAMSI	60
Db	337	TGSKNRKIALKLAIEFQTDOSGKIIVSTOEKELIVQPESSLFPRVEYIARAGWTRDDKYAMAM	366
Qy	61	ILDRSQTRQLQIVLISPELFIPEDDVWMEKORLIESVPDSTPLLIYEETTDIMINIDIF	120
Db	397	FUDRPQWIMQIVLILPPLALFIPSTENEBORLASARAVPNVOPUYVEEYVIMVYNVDIF	456
Qy	121	HVEPQSH-EEBIEFIIPASECKTGFMHLVYKITSILKESKYKSSGGLRAPASDFKPIKEI	179
Db	457	YFPQSEGEDELCFLRANCKTGCFCHLYKTVAILKSGCYDWSBPFSPOEDFCKPIKEI	516
Qy	180	ALISGMEVLIGHHSGIQVDYRRLVYVEGTRKDSLEHNLVYVSYVNPAGEVTRLITDGYIS	239
Db	517	ALTSGMEVLIGHHSGSIWNEETKLVLYQGRKDPLEHNLVYVSYVEAAGELVRLITTGFS	576
Qy	240	HSCCISQHCDFPISKYSNOKNPHCVSLKLSASPEDPFCCKTEPMATILDSAGLPTYTP	299
Db	577	HCSMSQANDMRYVSHSSVSTRPCNHYKLKSGPDDPLHKQPRFMASMEASCPDYP	636
Qy	300	PEIF\$ESTTGTFLYGLMLYKPHDLQPKKRYPTVLFIYSGPO-----	340
Db	637	PEIFHFHTSDVRLYGMITYKPHALQKGGHPVLVYVGGPQVQLVNNSPKIKYLRNLTL	696
Qy	341	-----	340
Db	697	ASLGVAVVVIDGRSGCORGLRFEGLKNQMGQVEIEDQVEGLQFVAEKYGFILDSRAIH	756
Qy	341	-----VALAGAPVTLMIFFPDGYTERWYNGHNDONOGUYGSVA	379
Db	757	GMSYGGFSLMGLIHKPOVFKYALAGAVTYMMAVIDGYTERYNDVDENNQHGIEASVA	816
Qy	380	MOAEKRPSEBPNLLILHGFUDENVAFAHSTILSLFLVYAKGPYDQIYPOERHSIRVPS	439
Db	817	LHVEKLPENPNRLLILHGFUDENVAFFPTNVLVSQILRACKRPYDQIYPIPRBRHSIRPES	876
Qy	440	GEHYELHLHLYOEUL	455
Db	877	GEHYEVTLLHPLQEVYL	892

```

RESULT 10
US-09-976-674-27
Sequence 27, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akimsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIORITY APPLICATION NUMBER: US 60/240,117
PRIORITY FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO:27
LENGTH: 892
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-674-27

```

Query Match	60.2%	Score 1494.5;	DB 10;	Length 892;
Best Local Similarity	50.5%	Pred. No. 2.7e-139;		

	Matches	281	Conservative	69	Mismatches	105	Indels	101	Gaps	2
Qy	1	TGTANPKYTFKMSSEIMIDAEGRIIDVIDKELIOPPEILFEGVEYIARAGMTPEBGKYANSI	60							
Db	337	TGSKNPKALKLAEEQOTDSOGKIVSTOEKELVQPPSSLFPPKVEYIARAGMTDGRYAMAM	396							
Qy	61	LLDRSQTRLQVILIPPELFIPEDDVMERORLIESVPSVTLIYEETTDIMINIDHF	120							
Db	397	FLDRQQLQVLLPFLPALFIPSTENBEQRLAARAVPRVQYVYEEVTNWINVHDF	456							
Qy	121	HVFPQSH-EEIEEPIFASBECKTGPHNLKYITSILKESKYKRSGLPADSDKCPICEI	179							
Db	457	YFPFQSEEDDELCEFRANECKTGFCMLYKVTAVLKSOGDWSBPSPGDEBEKCEPIKEE	516							
Qy	180	ALITGSEWVLGRHGSNIQVDEVRRLVYEGTSDSPLEHLVYVSVNPEEVRRLTDGRS	239							
Db	517	ALITGSEWVLARHGSKIVNEBETKLVPFGVTDTPHEHLVYVSTBAAGEIYRLTTPGFS	576							
Qy	240	HSCCISQCHDFPFIISKYSNOKNPHCVSLYGLSSPEDPTCKTKEFATIIIDSGAPLPDYR	299							
Db	577	HSCSNQSNFDMVSHYSSVSTPCVHVYTLSDPDDPLHKQRFNAMSMEAAASCPEDYR	636							
Qy	300	PEIFSEESTGTLTYGMLYKPHDLOPGKKYPRVLYTGGPO-----	340							
Db	637	PEIFHFHTRSDVRLGMLTYKPHALPGKKHPTVLVYGGPQVQVYVNSFKGIKYLRLMTL	696							
Qy	341	-----	340							
Db	697	ASLGAIVAVVIDRGSGCQGLRFEGLAKYKMGQVEIEDQYEGLOFAEKKGYFIDLSRVAIH	756							
Qy	341	-----VAIAGAPVTLMIFFDGTGTERYMGHPDONEGYLYGSVA	379							
Db	757	GWSYGGFLSLMGLIHKPQVFKVAILAGAPVTWMAADVTGTERYMYDVPPENNQHGYEAGSVA	816							
Qy	380	MOAEKFPSEPNLLHLHGFLDENVHPAHTSILSFLVRAKGRYDQIYQOEHSIRVPES	439							
Db	817	LHVEKLPEPNRLLHLHGFLDENVHFPHTNFLVSQILIRAKGYOLOIYPNEHHSIRCPBS	876							
Qy	440	GEHYELHLLHYLQENI-455								
Db	877	GEHYEVTLLHFLQEYL-892								
RESULT 11										
US-09-976-674-33										
; Sequence 33, Application US/09976674										
; Patent No. US20020115843A1										
; GENERAL INFORMATION:										
; APPLICANT: Oi, Steve										
; APPLICANT: Akimsanya, Karen										
; APPLICANT: Riviere, Pierre										
; APPLICANT: Junten, Jean-Louis										
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV										
; FILE REFERENCE: 70669										
; CURRENT APPLICATION NUMBER: US/09/976, 674										
; CURRENT FILING DATE: 2001-10-12										
; PRIOR APPLICATION NUMBER: US 60/240,117										
; PRIOR FILING DATE: 2000-10-12										
; NUMBER OF SEQ ID NOS: 61										
; SOFTWARE: PatentIn version 3.1										
; SEQ ID NO 33										
; LENGTH: 879										
; TYPE: PRT										

QY	61	LLDSSQRRLOLVLLSPELFIPEVDVMEKROLIESVPSVPTLLIYEETTDIWINHDI	120
Db	399	FLDRPQOWLQVLLPPLPFIISTEEBQRLASARVPRNVQVYVEETVWVINHDI	456
QY	121	HVFPDOSH-EEEIEFIIFASECKTGFRNHLKYKITSILKESYKKSGGCLPAPSDFCRPIKEBI	179
Db	457	YVPFGSDEDELCFLRANECKTGFCNLYKVTAVLKSQGYDMSEPPSPGDEDFKCPIKEBI	516
QY	180	AITGSEWVLGRHGSNIQVDEVRLVYFEGTKDSFLNHLVYVSVNPGCVTRLIDRGS	239
Db	517	ALTSGEWVLLARHGS-----KGRKPTLEHLLHYVSVSEAAAGEIYRLTTGPGS	563
QY	240	HSCCISQCDPEFISXSNQKNPHCVSLYKLSLSPEDDPCTCKTEPATIILDSAGPLPDYR	299
Db	564	HSCMSQNFDMFVSYHSVSTRPCVNHVKKLSGPDDDLHKQRFPMASHMAASCPDDYR	623
QY	300	PEIFSFSSTGFTLLGYMLKPHDLOPGKKYPTVLPFYGGPQ-----	340
Db	624	PEIFHFHTRSDVRLYGYMKYKPHALQPKKHPTVLFTVYGAPQVQLVNNSEFKIKYLRNLTL	683
QY	341	-----	340
Db	684	ASLGYAVVVIDGRSCQGRLEFEGALKNMQQVEIEDVEGLQVFAEKYGFIDLSRVAIH	743
QY	341	-----VAIAGAPVTLMIFYDTGTERBYMGHPDQNEQYVLGSVA	379
Db	744	GMSYCGFLSLMGLNGLIHKKPQVFKAVALNAGAVTVMMAVDTGTERBYMNVPENNQNGYEAGSVA	803
QY	380	MOAEKFPSEPNRLLIHGFIDBNVFAHTSILSLFVLPAKRPYDLOIYPOERHSIRVPES	439
Db	804	LHVELCLPNEPNRLLIHGFIDBNVAFPTNPLVSQLIRAKRPYQLOIYPNERHSIRCPES	863
QY	440	GEHYELHLHYLOENTL 455	
Db	864	GEHYEVTLLHFLQETYL 879	
RESULT 12			
US-09-976-674-35			
; Sequence 35, Application US/09976674			
; Patent No. US20020115843A1			
; GENERAL INFORMATION:			
; APPLICANT: Qi, Steve			
; APPLICANT: Akinsanya, Karen			
; APPLICANT: Riviere, Pierre			
; APPLICANT: Junien, Jean-Louis			
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLIV			
; FILE REFERENCE: 70669			
; CURRENT APPLICATION NUMBER: US/09/976,674			
; CURRENT FILING DATE: 2001-10-12			
; PRIOR APPLICATION NUMBER: US 60/240,117			
; PRIOR FILING DATE: 2000-10-12			
; NUMBER OF SEQ ID NOS: 61			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 35			
; LENGTH: 879			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-976-674-35			
Query Match 58.2%; Score 1445; DB 10; Length 879;			
Beat Local Similarity 49.3%; Pred. No. 2,3e-134;			
Matches 274; Conservative 67; Mismatches 101; Indels 114; Gaps 3			
QY	1	TGTANPKVTGKSEIMIDAEGRIIDVIDKEHLQPEILFEGVGYIARAGWTEGKYAMS	60
Db	337	TGSKNPKIALKLAEGQTSQCKIVSTQCKELVQPPSSLPFKVEYIARAGWTDGKYAMM	396
QY	61	LLDSSQRRLOLVLLSPELFIPEVDVMEKROLIESVPSVPTLLIYEETTDIWINHDI	120
Db	397	FLDRPQOWLQVLLPPLPFIISTEEBQRLASARVPRNVQVYVEETVWVINHDI	456

QY	121	HYFPQSH-EEETLEFIPASCKTGFRLLYKITSILKESKYRSGGGLPAPSDFCPCIKKEI	179
Db	457	YFPQSEGEDELICFLPLANECKTGFCGLLYKVTALVKSQGYDMSFPSPGDEFCPCIKKEI	516
QY	180	AITSGMEVLTGRHGSNIQVDEVRLLYFEGTKDPSLEHHLYVVSYYNPGCVRTLTRGRYS	239
Db	517	ALTSGEVETLVARHGS-----KGTQTPLEHHLYVVSYYEAAAGSIYRLTTPGFS	563
QY	240	HSCCISQHCDFPISKYSNQKNPHCVSLYKILSPSEDDPTCKTKEFMATILDSAGPLPDYTP	299
Db	564	HSCSMSONDMVSHSVSYSTPCVYVYKLSGDDPDLPHQPRFMASMMEAACCPDYVP	623
QY	300	PIIFSPSESTGTGLYQMLTKRPHDLOQKKRPTYLFIYGGPQ-----	340
Db	624	PEIFHFHTSDVRLYGMITKPHALQFQKHPYLVFYGGPQVQLVNNSPFGITRYLLRNLTL	663
QY	341	-----	340
Db	684	ASLGAVVVYIDGRGSCQGRLEFGALKNQMGVEIEDQVEGLQFVAKCYGFIIDLRSVAIH	743
QY	341	-----VAIAGAPLTIMFYDPTGTERYMGHPDQDEQGYILGSAVA	379
Db	744	GWSYGGFLSLMGLIHKRPQFKVAIIAAPATVMAAYDTGTERYMDVBPENNQHGYEAGSAVA	803
QY	380	MQAEKFPSEPNRLLILHGFLDENVHFAHTSILSLFVRAKGKPYDLOIYPOERHSIRVPES	439
Db	804	LHVEKLTPNPNRLLILHGFLDENVHFFHNLFLVLSQILIRBKGKPYQLQIYPERHSIRCPES	863
QY	440	GEHYELHLHYLOENL	455
Db	864	GEHYEVTLHLFLOEYL	879

QY	396	HGFLENNHFAHTSLLSLTVAGKPYDLOIYPORRHSIRVPESGGEYHLLHLYOENL	455
QY	211	HGFLENNHFAHTSLLSLTVAGKPYDLOIYPORRHSIRVPESGGEYHLLHLYOENL	300
Db			
QY	456	GSRIALAKVI	465
Db	301	GSRIALAKVI	310

```

RESULT 14
US-10-237-271-3
; Sequence 3, Application US/10237271
; Publication No. US20030096328A1
; GENERAL INFORMATION:
; APPLICANT: THE BURNHAM INSTITUTE
; APPLICANT: SMITH, Jeffrey W.
; APPLICANT: KIDDEL, Steven J.
; APPLICANT: AXELROD, Fumiko T.
; TITLE OF INVENTION: SERINE/THREONINE HYDROLASE PROTEINS AND SCREENING ASSAYS
; FILE REFERENCE: BURNL100-1
; CURRENT APPLICATION NUMBER: US/10/237,271
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 60/317,842
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-237-271-3

```

```

RESULT 13
US-09-993-959-4
; Sequence 4, Application US/09993959
; Publication No. US20030165489A1
; GENERAL INFORMATION:
; APPLICANT: Lacroix, Eric
; APPLICANT: Grouzmann, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/993,959
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatencIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-959-4

Query Match          55.5%; Score 1377.5; DB 12; Length 310;
Best Local Similarity 83.5%; Pred. No. 2,3e-128;
Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1

QY      207  FEFGTDSLEHLLVYVSVNBSGVYRLTRDRGSHSCSCCISOHCDPFIISKSNOKNPHCVSL 266
Db       1  FEFGTDSLEHLLVYVSVNBSGVYRLTRDRGSHSCSCCISOHCDPFIISKSNOKNPHCVSL 60

QY      267  YKLSPEDDPCKTKEFPATILDSAGPLPDYVPPELFPSESTTGFTLYGMLYKPHDLOPG 326
Db       61  YKLSPEDDPCKTKEFPATILDSAGPLPDYVPPELFPSESTTGFTLYGMLYKPHDLOPG 120

QY      327  KKYPYLVLEIYGGPO----- 340
Db       121  KKYPYLVLEIYGGPOGQIEIDDPVEGLQYLASRYDFIDLDRVGIHGMSYGYLSLMAAMQR 180

QY      341  -----VALAGAPYTLMTIYFDYDGYTTRWYGHPRQONQGYLLGSVMAQAERFSEPNRLLIL 395
Db       181  SDIFRVALAGAPYTLMTIYFDYDGYTTRWYGHPRQONQGYLLGSVMAQAERFSEPNRLLIL 240

```

Query Match	47.8%	Score 1187.5	DB 15	Length 508
Best Local Similarity	44.4%	Pred. No. 4e-109		
Matches 241	Conservative 64	Mismatches 93	Indels 145	Gaps 7
QY	23	IIDVIDKELIOPFELLPEGVEXIARAGCTPEGKXVMSILLDSQTRLQVLLSPLELFIPV	82	
DB	1	IVSQNKEKLVFPSSLFPKVEIYARAG-----AMAFELDPQQMLQVLLPALPFIIS	53	
QY	83	EDDMERORLLIESVPDSVTPLLIYEETDIIWINHIDIFHVPQSH-EEBIEFIIPASECKT	141	
DB	54	TENEBORLASARAAPRNQVPYVVAEEVNWMINVHIDIFYPFQ-EGEDELCEFLANECKT	113	
QY	142	GFRHLXKTSLLSKSKYRRSSGGIIPAPDFKCIPEELTAITGCEHVLGRHSNIOYDEV	201	
DB	114	GFCHLYKTAALKKQGYMSEPSFG-----EGEGLTNAIW-----VNEE	154	
QY	202	RRLVFEETKDSPLEHLLYVSVYVPGSEVTRLTDGRGSHSCCISOHCDFPISKYSNQKP	261	
DB	155	TKLVYFGQCTKPLREHLLYVSVYEAAGIYKLTITGFCFHSQSMQFDMFVHYSVSIRP	214	
QY	262	HCVSLYKLSSEDDPTCKTEKFMATILDSAGRLPDYTPPEIIFSFESTGTFLYGLYKPH	321	
DB	215	PCVHVYKLSGDDDDLHKQPRFMSMMEA-----KIFPHFRSDVRLYGLMYKPH	265	
QY	322	DLQGGKYPTLFIYGGPO-----	340	
DB	266	ALQGGKGRPTLVFYGGPOVQLVNNSPKGIKYLRLNTLASLGAYAVVVIDGRGSCQGRLEF	325	
QY	341	-----V	341	
DB	326	EGALKNQMGVIEIDQVEGLQPAVAKCYFIDLSPRAIHGMSYGGELSLMGLHKQVFEV	385	
QY	342	AIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAMQAEKPPSEPNRLLHLGFLDE	401	
DB	386	AIAGAPVTVMMAUYDTGYTERYMDVPENNQHGYEAGSVALLHEKLPNEPNRLLILHGFLDE	445	
QY	402	NMHAFHSLISLFLVRGKPYDL-----QIYPOHRHSIRVSESEHVELHLLHYLO	452	
DB	446	NMHFPHNPLFVSLIRAGKPYOLOVALPVPSPQIYPMERHSIRCPSESEHVEVTLHLFLO	505	
QY	453	ENTL 455		

Db 506 EYL 508

RESULT 15  
US-09-976-674-29  
; Sequence 29, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akinsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29  
; LENGTH: 832  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-674-29

Query Match 44.1%; Score 1094.5; DB 10; Length 832;  
Best Local Similarity 57.3%; Pred. No. 1.5e-99;  
Matches 197; Conservative 58; Mismatches 88; Indels 1; Gaps 1;  
QY 1 TGTANPKVTFKMSSEIMIDAEGR1IDVIDKELIQFEILFEGVEYIARAGWTPEGKYAWSI 60  
DB 337 TGSKNPKRIAKLAEPFOTDSQKIVSTQEKELVQPFSSLPKVEYIARAGWTRDGKYAMAM 396  
QY 61 LDRSQTRLOIVLISPELFIPEDDVMERQRLIESVDSVTPLIYBETTDIMINIHDI 120  
DB 397 FLDRPOQWLOLVLLPPLFLPSTENEQRLASARAVPRNVQPYVVEEVTNWVWVHDIF 456  
QY 121 HVPFOSH-EEIEFIFASECKTGFRHLKITSILKESKYRSSGGLPAPSDFKPIKEEI 179  
DB 457 YFPQSEGEDDELCTFLANECKTGFCFLYKTAVALKSQGYDMSSEPFSGEDEFKCPIKEEI 516  
QY 180 AITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVSYVNGEVRRLTDGYS 239  
DB 517 ALTSGEWEVLARHGSKIWNVEETKLVEYFGTKDTPLEHLLVYVSYENAGEIVRLTRGFS 576  
QY 240 HSCCIGQHCDPFIKSYNSQNPCHCVSLYKLSPEDDPTCTKEFMATILDSAGPLPDYTP 299  
DB 577 HSCSMQNFDMFVSHSVSVSTPPCVHYKLSGPDDELHAKQPRFWMASMMEAASCPDYP 636  
QY 300 PEIFSFEFTGTFLYGLLYKPHDLQPGKKYPTVLFYGGPOVAI 343  
DB 637 PEIFHFHTRSDVRLYGMITKPHALQPGKKHPTVLFYGGPOVL 680

Search completed: October 15, 2003, 17:28:51  
Job time : 67.704 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 12.2186 Seconds  
(without alignments)  
3659.853 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKYTFKMSKIMIDAE.....HLHLVLENTGSRALAKVI 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	462	18.6	931	2	TJ32919
2	429.5	17.3	738	2	A87516
3	355	14.3	741	2	JCS142
4	300.5	12.1	711	2	S66261
5	277	11.2	818	1	A30107
6	267.5	10.8	931	2	A49737
7	267	10.8	793	2	T41703
8	248.5	10.0	792	1	A39914
9	238	9.6	766	1	CDH26
10	230	9.3	760	1	S23752
11	226	9.1	779	2	T25173
12	226	9.1	799	2	T25174
13	221	8.9	803	2	A41793
14	216	8.7	803	2	T68600
15	216	8.7	865	2	T54331
16	176	7.1	829	2	T19514
17	173.5	7.0	795	2	F82858
18	160.5	6.5	759	2	T38593
19	145.5	5.9	743	2	T37700
20	139	5.6	642	2	C71137
21	136.5	5.5	709	2	B82580
22	134	5.4	683	2	H75007
23	131	5.3	631	2	H75007
24	113.5	4.6	466	2	B89877
25	111.5	4.5	1018	1	CGH2A
26	108.5	4.4	1415	2	S52267
27	105	4.2	692	2	S61200
28	104.5	4.2	265	2	B84063
29	104	4.2	1438	2	C89900

30	103.5	4.2	1835	2	S46082	urea carboxylase (
31	103.5	4.2	4452	1	YGSSG2	gramicidin S synth
32	102.5	4.1	592	2	C70108	peptidase homolog
33	102	4.1	732	1	A35655	peptidyl-dipeptida
34	102	4.1	1312	1	A34171	peptidyl-dipeptida
35	101.5	4.1	4450	2	UX0340	gramicidin S synth
36	901	4.1	1313	1	JC2038	peptidyl-dipeptida
37	100	4.0	558	2	A43667	serine C-palmitoyl
38	99.5	4.0	543	2	S46098	probable acid-CoA
39	99.5	4.0	591	2	H72474	probable acylamino
40	99.5	4.0	1948	2	B69511	N conserved hypoch
41	97	3.9	866	2	T79267	trial protein - Bac
42	96.5	3.9	440	2	T24073	hypothetical prote
43	96.5	3.9	957	2	H68141	hypothetical prote
44	96.5	3.9	2560	1	I40457	peptide synthetase
45	96	3.9	413	2	D72260	aminotransferase,

ALIGNMENTS

RESULT 1									
TJ32919									
hypothetical protein K02F2.1 - Caenorhabditis elegans									
C:Species: Caenorhabditis elegans									
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999									
C:Accession: TJ32919									
R:Maggi, L.; Goela, D.									
submitted to the EMBL Data Library, January 1998									
A:Description: The sequence of C. elegans comid K02F2.									
A:Reference number: 221246									
A:Accession: TJ32919									
A:Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-931 <MAG>									
A:Cross-references: EMBL:AF043699; PDB:AA897564.1; GSPDB:GN00019; CESP:K02F2.1									
A:Experimental source: strain Bristol N2; clone K02F2									
A:Genetics:									
A:Gene: CESP:K02F2.1									
A:Map position: 1									
A:Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2; 5									
Query Match									
Best Local Similarity 18.6%; Score 462; DB 2; Length 931;									
Matches 954; Conservative 79; Mismatches 177; Indels 192; Gaps 23;									
QY	2	GTANPKYTFKMSKIMIDAEGRIVDIDKELIQPFELFEG--VEYIARAGWTREGKVM	58						
DB	355	GTGNAYSTLRM---VILENGKAYDVPKLD-----EVIYKCPFEYITRAGFFSDGTTW	406						
QY	59	SILDRSQRLQIVLI-----SPELFIVEDDWMERORLIESV	96						
DB	466	PRGKLRGVQIHKARNDDYMINHNAIYPLKTTDEBHPWBEFYCLEKPNQ--SCLLLISAE	524						
QY	97	PDSVT--PLIYEETDWINIHDFHVPQSHSE--IEFFASECKTGPHLYKITSI	152						
DB	407	VQVMSRDOACSILLIPYTDILLPELGSIKEDNLQSTDLNMGWMD--KSHBEETMR	465						
QY	153	LKESKYKSSGGLPAPSDFKCPIKEIATITSGEVLGRHGSNIQVDEVRALVVEGTGD	212						
DB	525	LDQNY-----CRHTEKILMENFET--NKSNGIYVDEVRRELYVYANES	568						
QY	213	SPLBHLVYVYVNGEYTRLDNGYSHSCCISQHCDFPISRYSNQK---NPHCVSLYK	268						
DB	569	HPTENNI--CVSHYRTGHAQLTESGI-----C--FKSBRANGKLLALDDHGFACVM	616						
QY	269	LS--SPEDDPCKTKEF--W-----ATILDSAGP--LPD--YTPPEIFSES--TT	309						
DB	617	TSVGSFAE---CRYSFRKENEVLPTVYANITVSGHPQPDHAFSPFEMIEFQSKKT	673						
QY	310	GFTLYGMLYKPHDLQPKKYPFTVLVIYGP-----	339						
DB	674	GLMHYAMILARSPNDPYYKYPVPHYVGGPQIQIVHNFQSWIQTIRFCRLGVVVFIDNR	733						

QY 340 ----- 339  
Db 734 GSAHRIEPRRIHKKMGIVVEDQVEGLMLAERTGPFMDNRVVVHGMVYGYMALQM 793  
QY 340 -----OVALAGAPYTLMIFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391  
Db 794 IAKHPNIYRAALIGGAVSDMRLYDTATYTERYMGVP--LEEHVYGASSITGLVEXLPDEPNR 852  
QY 392 LILLHGFLENVHFAHTSILSLFLVRAGKRPYDLOIYPOEHSIRVPESGHEYLHLHL 451  
Db 853 LMLVHGLMDENVHFAHLHLVLDRCIKKKMHELVIFPNEHGVNRNDASITYLDARMYFA 912  
QY 452 OE 453  
Db 913 QQ 914  
RESULT 2  
A87516  
dipeptidyl peptidase IV (imported) - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: A87516  
R:Netman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolot  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: A87516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-738 <SNO>  
A:Cross-references: GB:AE005673; NID:gl3423647; PIDN:AAK24125.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC2154  
Query Match 17.3%; Score 429.5; DB 2; Length 738;  
Best Local Similarity 24.9%; Pred. No. 3.6e-26;  
Matches 123; Conservative 57; Mismatches 146; Indels 167; Gaps 12;  
QY 44 YVARAGTPEGKXAMSLDRSQRLQIVLISPELFIPEVDVMEKORLIESVDSYTPL 103  
Db 284 YVARVAMSADKTVYVQLSRDQKTLDLAF-----DATGA 320  
QY 104 --IIVEETDIWNIHDIPIVPPQSHBEIEFIASECTGPHLYKITSILKSKYKRS 161  
Db 321 GKTILLDTDPHFLEVSNDPRLTDG-----TFLKGE-KDGNQHLRYA-----A 364  
QY 162 SGGLPAPSPDFKCPKEIEIAITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYV 221  
Db 365 DGLILA-----QITKGMPIVIGLEG-----VDEARKVAIFASIDTPIERLYE 408  
QY 222 VSVVNGEVRLTLDRGYSHCISQHCDFISKYSNQKNPCHCVSLYLSSP-----EDDP 276  
Db 409 VSYAKGKPKRALTSAGGWMVAKVADNGAFAGTSDKTPSQALYSADSKRVRMIENK 468  
QY 277 TCKTEKFWATILDSAGPLPYTPPELIFSFESTTGFTLYGMLVYPHDLQPKKYPVLYFIY 336  
Db 469 LAEGHPYV---PYANLPO---PEFOSLKAADGETLHTEILKPIGDFPAKKIPALYSVY 521  
QY 337 GGP----- 339  
Db 522 GGPHAQRVMKNMHSPEPTYLEAGYVIFKLDNRGSGNRSAKFMRALDRKLGTVEVEDQL 581  
QY 340 -----OVALAGAPYTLMIFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391  
Db 582 GAKFLASQPYVDADKLGVGMWSYGGFAMALLTAENTPFRAAGAAPTWSLYDTATYTE 641  
QY 361 RYMGHPDNEGGYLLGVAMQAEKFPSEPNRLILLHGFLENVHFAHTSILSLFLVRAGK 420  
Db 642 RYMGKBDENAGAYASDINNRIKTL--APQSILLHGMADDNVIFENSTRLMALQKRAI 639

QY 421 PYDLOIYPOEHS 433  
Db 700 LPEMAMYPERHS 712  
RESULT 3  
JC5142  
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia  
C:Species: Xanthomonas maltophilia  
C>Date: 31-Jan-1997 #sequence\_revision 27-Feb-1997 #text\_change 20-Jun-2000  
C:Accession: JC5142  
R:Kashima, T.; Ito, K.; Yoshimoto, T.  
J. Biochem. 120, 1111-1117, 1996  
A:Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expressic  
A:Reference number: JC5142; MUID:97164011; PMID:9010758  
A:Accession: JC5142  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-741 <XAB>  
A:Cross-references: DDBJ:DB3263; NID:g1753196; PIDN:BA11872.1; PID:g1753197  
C:Comment: This enzyme catalyses the hydrolytic removal of amino-terminal dipeptidyl re  
C:Keywords: dipeptidyl-peptidase IV  
F:4-18/Domain: transmembrane #status predicted <Trm>  
F:610/Active site: Ser #status predicted  
F:685/Active site: Asp #status predicted  
F:717/Active site: His #status predicted  
Query Match 14.3%; Score 355; DB 2; Length 741;  
Best Local Similarity 22.2%; Pred. No. 3.3e-20;  
Matches 123; Conservative 77; Mismatches 164; Indels 190; Gaps 17;  
QY 2 GTANPYTPFMSRIMIDAEGRITVDIKELIQFELLBEVEYIARAGWTPEGKXAMSL 61  
Db 257 GDANVQVQLGVSPADQQTWIDLGKQDT-----VLARNMDDPQHLFQ-R 304  
QY 62 LDRSQRLQIVLISPELFIPEVDVMEKORLIESVDSYTPLIIVEETDIWNIHDIH 121  
Db 305 QSDQDKKLDVETL-----ASNQR-----VLAHETSTWPLHNSLR 343  
QY 122 VPPQSHBEIEFIPEASECTGPHLYKITSILKSKYKSSGGLPAPSPDFKCPKEIEIAI 161  
Db 344 FL-----DQSLIMSSB-RTPQHLYRI-----DSKGA- 374  
QY 182 TSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVSVNGEVRLTLDRGYS 241  
Db 375 THGNMSV---DELLAVDEKAGLAYFRAGIESAREQIYAVP--LOGGQPORLKAQGMHS 429  
QY 242 CCISQHCDFEISKYSNQKNPCHCVSLYLSSPEDDPTCKTEFWATIL--DSAGPLPYT- 298  
Db 430 ASPARNASVYVDSWNSNSTPQIEPFRANG-----EKIATLVENDLADPKHPYAR 479  
QY 299 -----PPELIFSFESTTGFT-LYGMLYKPHDLQPKKYPVLYFIYGP----- 339  
Db 480 YREAPRPVEGTLTAADGKTPLYSVYKAGFPDARKYPAVAVVYVGGPASQVTDSWGR 539  
QY 340 -----OVALAGAPYTLMIFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391  
Db 540 GDHLFNYQLAQQSYVFSLDNRGTPRRGRDFGALYKGGKTVAVLAQLRGVAMLKQOPWV 599  
QY 343 -----IAGAPYTLMIFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391  
Db 600 DPARIQVQSGWSNGCYTLMMLAKASDSYACGVAGAEVTDWGLVDSHYTERYMDLPARND 659  
QY 372 GYLLGVAMQAEKFPSEPNRLILLHGFLENVHFAHTSILSLFLVRAGKRYDQIYPOER 431  
Db 660 GYREARVLTHIEELRSP---LILHGMADDNVIFENSTRLMALQKRAI 639  
QY 432 HSIRVPESGHEYL 445  
Db 717 HGLSGADALHRYRV 730



A:Residues: 1-931 <SN>  
A:Cross-references: GB:L21944; NID:g347196; PIDN:AAA35119.1; PID:g3477197  
R:Anna-Artiola, S.S.; Herskowitz, I.  
Yeast 10, 801-810, 1994  
A:Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase  
A:Reference number: S45451; MUID:9506382; PMID:7975897  
A:Accession: S45451  
A:Molecule type: DNA  
A:Residues: 1-931 <ANN>  
A:Cross-references: EMBL:L21944; NID:g347196; PIDN:AAA35119.1; PID:g347197  
R:Gallison, F.; Dujon, B.  
Submitted to the EMBL Data Library, October 1995  
A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome  
A:Reference number: S60938  
A:Molecule type: S60946  
A:Molecule type: DNA  
A:Residues: 1-931 <GAL>  
A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771  
R:Boyer, J.; Fairhead, C.; Gallion, L.; Gallison, F.; Michaux, G.; Thierry, A.; Dujon, B.  
Submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67104  
A:Accession: S67112  
A:Molecule type: DNA  
A:Residues: 1-931 <BOY>  
A:Cross-references: EMBL:Z75127; NID:g1420507; PIDN:CAA99437.1; PID:e252398; PID:g1420507  
A:Experimental source: Strain S288C  
R:Gallison, F.; Dujon, B.  
Yeast 12, 877-885, 1996  
A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV  
A:Reference number: S71713; MUID:96437977; PMID:8840505  
A:Accession: S71721  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-931 <GMB>  
A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771  
A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
A:Genetic: SCD:STE13; YC11  
A:Cross-references: SCD:S0005745; MIPS:YOR219C  
A:Map position: 15R  
C:Function:  
C:Description: involved in processing of alpha-factor prepropheromone  
C:Keywords: dipeptidylpeptidase hydrolase; glycoprotein; transmembrane protein  
F:113-150/omam: transmembrane #status predicted <TM>  
F:377/Binding site: carbonylase (asn) (covalent) #status predicted  
F:785,863/Active site: Ser, Asp, His #status predicted  
Query Match 10.8%; Score 267.5; DB 2; Length 931;  
Best Local Similarity 21.3%; Pred. No. 4.7e-13;  
Matches 114; Conservative 63; Mismatches 164; Indels 195; Gaps 18;  
OY 74 ISPELFLPVEDDWERORLIESV---DSVPLIIEETD---WI-NIHDIHFVFPQ 125  
DB 445 ISPTTF---RPEIRDRNSKILIDVKYDIPSSQMLTVNTNSNLFNGMIETKTKILSLSPK 501  
OY 126 SHEEIE--FIFASECKTGFRHLKYTKISILKESKYKSSGGLPAPSDPKPIKEIAYTS 183  
DB 502 PELKRMVGYGIDIHADSRGFSHLFYPIV-----AKEPIOLTK 540  
OY 184 GEMEWLGRH--GSNIQVDEVARLVFEGCTKDSPLEHLLYVVSYY-----NPG 229  
DB 541 GNMENVGTGNGIAGVEYETD---TIFPANEIGVMSQHLYSISLSDSTTQMTFOSLQNP-- 594  
OY 230 VTRLTIDGYSHSCCISQCHDFIFISKYNSQNP-----HCVSLYKLSPE 273  
DB 595 ---SDKYDFDPFLSSARAYASKKGLPDTPIKVAQPLTRVLVAIEIHDSILQTKDE 650  
OY 274 DDPECKTKEFWATILDSAGPLPDYTPPEIFSEFST---GFTLYGMLYKKPHDLOPKKYP 330  
DB 651 ---KFKK---KIKNYDLP-ITSYKTMVDDGVEINVIYIIRKANLNPKKKYP 694  
OY 331 TVLFYIGP----- 339

```

Db      695 ILNNVGGPGQGTFTTKSLAFEGAAVWSGLDVIVLQLEPRGTGGKGMFSRMAEKLGYM 754
QY      340 -----QVALGAPRTL 350
Db      755 EPRDITEVTKFPIQRNSQIHDESKIAIWMGSYGGFTSLKTVELDNGDTFKYMAVAAPVTN 814
QY      351 WIFPDVTGTERYMGHPDPONEGYYIGSVAAOAEKFPSEPNRLLLHGFGLDENVHFAHTSI 410
Db      815 WTLVDSVYTERYNNQPSNEHGTYEVSTIQKFKFES-LKRLFLVHGTFDDNVHIGNFR 873
QY      411 LLSFLVRAG-KPYDLQIYPOERHSIRVPESGEHYELHLHYLOENLGSRIALKYI 465
Db      874 LVQDLNLGLTNYDMHIFPDSDSHSIRYHNAQRIYFOKLYWLRLDAFAERPNTEVL 929

RESULT 7
T11703
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
C|Species: Schizosaccharomyces pombe
C|Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C|Accession: T11703
R|Murphy, L.G. Harris, D.; Bartell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1998
A|Reference number: Z22011
A|Accession: T11703
A|Status: Preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-793 <MUR>
A|Cross-references: EMBL:AL031180; PINN:CAA20138.1; GSPDB:GN00066
A|Experimental source: strain 972h-, cosmid c2E11 -chimeric
C|Genetics:
A|Gene: SPAC2E11.08
A|Map position: 1
C|Superfamily: dipeptidyl-peptidase IV

Query Match      10.8%; Score 267; DB 2; Length 793;
Best Local Similarity 25.6%; Pred. No. 4,16-13;
Matches 100; Conservative 37; Mismatch 127; Indels 126; Gaps 14;

QY      179 IAIISGEWEVLGRHGSNIQVDEVSLVYFEGTKDSPLEHILYVSVNPGSEVRLTRGRY 238
Db      412 IYLTSGAMDVT---DGFHIDGDFGNVYFLATLKDSTERHLYVS-LDTLEIYGITDNGE 467
QY      239 S--HSCCISQHCDFPISKYSNQKNPVCVLSLYKSSPEDDFTCKTKEFMATILDSAGPLP 295
Db      468 DEGYSTSFSPFGPFYLYNHGPRVP-----WQELRSTKDXCYCSLENSRLKQO---LS 520
QY      296 DYTPPEI---FSPESTTFTGLYGLKRPDLQGGKKYPTVLFYGGP---QVA----- 342
Db      521 SITLPSEYGLKLTNDTT-FNF-MERRPRNFVNKKYPLVFFAYGGPGSQVAKLPRVD 577
QY      343 ----- 342
Db      578 FOATLASHPDPFIVVTLDRGTGTFGNAPFVSVSRHLGSEWSDGQAGKFMADLPFVD 637
QY      343 -----IAGAVTLMIFPDGTGRYMGHPDPOEGY 373
Db      638 ENHVGIGMSYGYLTLLKTLTQDVESYGMAVAAPVLTWRLDYSYTERYMDLPQYNKEG- 696
QY      374 YLGSVAAOAEKFPSEPNRLLLHGFGLDENVHFAHTSILLSFLVRAG-KPYDLQIYPOERH 432
Db      697 YKNSQIHDEKF-KQLKRFPVAHOTGDVNHFGQSHMLMGMLNLCANCYNDMAVFPDSAH 755
QY      433 SIRVPESGEHYELHLHYLOENLGSRIAL 462
Db      756 SI-----SYHNASLSIYHRLSEWIGDALGRI 781

RESULT 8
A339914
dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
N|Alternate names: GP10; membrane glycoprotein 110K; OX-61
N|Contains: dipeptidyl-peptidase IV, soluble form

```



C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999  
 C:Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781  
 R:Hong, W.; Doyle, D.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987  
 A:Title: cDNA cloning for a bile canalliculus domain-specific membrane glycoprotein of rat  
 A:Reference number: A39914; MUID:88068516; PMID:3479775  
 A:Accession: A39914  
 A:Molecule type: mRNA  
 A:Residues: 1-792 <HON>  
 A:Cross-references: GB:J02997; NID:9204463; PIDN:AAA1272.1; PID:9204464  
 R:Ogata, S.; Misumi, Y.; Ikehara, Y.  
 J. Biol. Chem. 264, 3596-3601, 1989  
 A:Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA  
 A:Reference number: A33315; MUID:89123496; PMID:2563382  
 A:Accession: A33315  
 A:Molecule type: mRNA  
 A:Residues: 1-37, 'A', 39-182, 'I', 184-331, 'T', 333-351, 'C', 353-393, 'W', 395-561, 'L', 563-623,  
 A:Cross-references: GB:J04591; NID:9203973; PIDN:AAA1096.1; PID:9203974  
 A:Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332  
 A:Accession: B33315  
 A:Molecule type: protein  
 A:Residues: 1-20, 35-54, 427-443, 505-509, 511-520, 530-538, 593-600, 602-608, 618-627 <H02>  
 R:McCaughan, G.W.; Wickson, J.E.; Creswick, P.F.; Gorrell, M.D.  
 Hepatology 11, 534-544, 1990  
 A:Title: Identification of the bile canallicular cell surface molecule GP110 as the ectop  
 quence.  
 A:Reference number: A60730; MUID:90228896; PMID:1970322  
 A:Accession: A60730  
 A:Molecule type: protein  
 A:Residues: 28-47, 'XX', 50-53, 55-58 <MC>  
 R:Ogata, S.; Misumi, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.  
 Biochemistry 31, 2582-2587, 1992  
 A:Title: Identification of the active site residues in dipeptidyl peptidase IV by affinity  
 A:Reference number: A42203; MUID:92190188; PMID:1347701  
 A:Accession: A42203  
 A:Molecule type: protein  
 A:Residues: 'R', 625-630, 'X', 632-648 <OG2>  
 R:Iwaki-Bgawa, S.; Watanabe, Y.; Fujimoto, Y.  
 Biol. Chem. Hoppe-Seyler 374, 973-975, 1993  
 A:Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl p  
 A:Reference number: S38949; MUID:94128239; PMID:7905271  
 A:Accession: S38949  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Status: preliminary  
 A:Residues: 281-302 <IMA>  
 R:Hong, W.; Doyle, D.  
 J. Biol. Chem. 263, 16892-16898, 1988  
 A:Title: Membrane orientation of rat GP110 as studied by in vitro translation.  
 A:Reference number: A31781; MUID:89034185; PMID:3182821  
 A:Accession: A31781  
 A:Molecule type: mRNA  
 A:Residues: 1-40 <H03>  
 C:Comment: This protein is localized to the bile canalliculus, which is the apical domain  
 C:Superfamily: dipeptidyl-peptidase IV  
 C:Keywords: dipeptidyl-peptidase hydrolase; glycoprotein; homodimer; liver; serine protease  
 F:1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATM>  
 F:1-38/Domain: signal sequence #link MATS #status experimental <SIG>  
 F:1-6/Domain: intracellular #status predicted <INT>  
 F:7-8/Domain: transmembrane #status predicted <TMN>  
 F:29-34/Domain: extracellular #status predicted <EXT>  
 F:35-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATM>  
 F:83-90, 148, 217, 227, 319, 521, 686/Binding site: carbohydrate (Aam) (covalent) #status pred  
 F:631/Active site: Ser #status experimental  
 F:709, 741/Active site: Asp, His #status predicted  
 Query Match 10.0%; Score 248.5; DB 1; Length 792;  
 Best Local Similarity 21.0%; Pred. No. 1.2e-11;  
 Matches 122; Conservative 66; Mismatches 188; Indels 205; Gaps 23;  
 2 GTAPKVTFFKSEIMIDAGRIIDVIDELIOPFILTBECHYIARACGTBEGKAWSEIL 61

Db 258 GAVNPFVKPF--VNTDSLSTTTTIPMQITAPASVT-TGDHYLCDVAVWSEDRIS----- 310  
 Qy 62 LDRSGTRQIVLISBELFIPVEDDMERORLLESPPDSTPLIYEETDININHDIFH 121  
 Db 311 -----LQWLRRIQNY--SYMAICDDVKTNLWNPCPTQEH 343  
 Qy 122 V-----FPOSHEIEFIFASECKTGRHLKYKTSILKESKYRSSGGLPA 167  
 Db 344 IETSAITGVNRRPRAEPHFTSDGSSFYKIVSKDQYK---ICQKQDK----- 390  
 Qy 168 PSDFKPIKEIAITSGEMVYGRHGSNIQVDEVRLVYFEET-KDSPLEHLLVYVSYVN 226  
 Db 391 -----PEODCTFIRKAMEVI-----SIEALTSPLYIYSNEYKEMPGGRNLKYI----- 435  
 Qy 227 PEGVRLTRDRGSHSCCIS-----QHCFP---ISKYSNQKPKC-----VSLYKLSPPD 274  
 Db 436 -----QLTD--HTNKKCLSCDLNPRCCOYVSLSKEKAYQAGRGPLPLTYLHRSTD 488  
 Qy 275 DPTCKTEFMAITLDSAGPLPYTPPE---LPSFSTTGFTLYGMLYKPHDLPQCKKPT 331  
 Db 489 QKEALVLED-NSALDKM--LQDVQMPKSLDFIVLNTRF-WYQMLPRH-FDKSKKPL 543  
 Qy 332 VLFYIGP----- 339  
 Db 544 LIDVAGFCOSKADAPFPNMTYLASTENITVASFDGSGSYQGDKIMHAIKRLGTLE 603  
 Qy 340 -----QVAIG-----APVTLWTFY 354  
 Db 604 VEDQEARQFLKMGFVDSKQVAILWMSYGYVTSMLVSGSGVFKCGIAPVPSRMEY 663  
 Qy 355 DPTGYERYMG--HPDQNGGYVLGSVMAQKPEPRLILLHGFLENVFAHTSILL 412  
 Db 664 DSVYTERYMGLEPTEDNLDHRTNSTVMSRAENF--KQVEYLITHTADNVHFOQSAQIS 721  
 Qy 413 SFLVAGKPYDQLQYPOERHSIRPESEGHVELHLHLYOE 453  
 Db 722 KALVDAGVDFQGMWYTDHDHGIASTAGHITYSHMSHFLOQ 762  
 RESULT 9  
 CDH026  
 dipeptidyl-peptidase IV (EC 3.4.14.5) - human  
 N:Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence revision 23-Aug-1996 #text change 18-Jun-1999  
 C:Accession: S24313; B42408; B61136; S59510; S16154; S59857; S15520  
 R:Misumi, Y.; Hayaishi, Y.; Arakawa, F.; Ikehara, Y.  
 Biochim. Biophys. Acta 1131, 333-336, 1992  
 A:Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a se  
 A:Reference number: S24313; MUID:92329551; PMID:1352704  
 A:Accession: S24313  
 A:Molecule type: mRNA  
 A:Residues: 1-6, 'I', 8-766 <MIS>  
 R:Darmoul, D.; Lacasa, M.; Baricault, L.; Marguet, D.; Saphin, C.; Trotot, P.; Barbat, A  
 J. Biol. Chem. 267, 4824-4833, 1992  
 A:Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cance  
 IV mRNA levels during cell differentiation.  
 A:Reference number: A42408; MUID:92165847; PMID:1347043  
 A:Accession: B42408  
 A:Molecule type: mRNA  
 A:Residues: 1-5, 'R', 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-766 <DAR1>  
 A:Cross-references: GB:M00536; NID:9181569; PIDN:AAA52308.1; PID:9181570  
 A:Experimental source: intestine  
 A:Note: this sequence corresponds with the author's translation  
 A:Accession: A42408  
 A:Molecule type: mRNA  
 A:Residues: 1-5, 'R', 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-766 <DAR2>  
 A:Cross-references: GB:M00536; NID:9181569  
 A:Note: sequence extracted from NCBI backbone (NCBI:83986, NCBI:83988); this sequence  
 R:Gorrell, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
 Gastroenterology 101, 618-625, 1991  
 A:Title: Expression of sucrase-isomaltase and dipeptidyl-peptidase IV in human small int

A:Reference number: A61136; MUID:91317403; PMID:1677636  
 A:Accession: B61136  
 A:Molecule type: protein  
 A:Residues: 1-15, 'X', 17-22 <GOR>  
 R:Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.  
 Biochem. J. 311, 835-843, 1995  
 A:Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a  
 A:Reference number: S59510; MUID:96067599; PMID:7487939  
 A:Accession: S59510  
 A:Molecule type: DNA  
 A:Residues: 1-31 <BOB>  
 A:Cross-references: GB:S79876; NID:g1195574; PIDN:AA835614.1; PID:g1195575  
 R:Tanaka, T.; Camerini, D.; Seed, B.; Totimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg, H.  
 J. Immunol. 149, 481-486, 1992  
 A:Title: Cloning and functional expression of the T cell activation antigen CD26.  
 A:Reference number: I56154; MUID:92325476; PMID:1352550  
 A:Accession: I56154  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-436, 'S', 438-766 <TN>  
 A:Cross-references: GB:M74777; NID:g180082; PIDN:AA51943.1; PID:g180083  
 R:Abbot, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.  
 Immunogenetics 40, 331-338, 1994  
 A:Title: Genomic organization, exact localization, and tissue expression of the human CD  
 A:Reference number: S59857; MUID:95012454; PMID:7927537  
 A:Accession: S59857  
 A:Molecule type: DNA  
 A:Residues: 1-436, 'S', 438-766 <ABB>  
 A:Cross-references: EMBL:U13734  
 C:Genetics:  
 A:Gene: GDB:DP4  
 A:Cross-references: GDB:125239; OMIM:102720  
 A:Map position: 2q24.3-2q24.3  
 A:Insertions: 2/3; 32/1; 65/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3  
 F:7-28/Domain: transmembrane #status predicted <TN>  
 F:29-766/Domain: extracellular #status predicted <EXT>  
 F:85,92,150,219,229,281,321,950,665/binding site: carboxydrate (Asn) (covalent) #status  
 F:630,708,740/Active site: Ser, Asp, His #status predicted  
 Query Match 9.6%; Score 238; DB 1; Length 766;  
 Best Local Similarity 21.1%; Pred. No. 8.3e-11;  
 Matches 120; Conservative 71; Mismatches 193; Indels 186; Gaps 24;  
 Db 2 GTANPVTPEKMSIMIDAEGRITDVIDKELIOPELIFEGVEYIARAGWTPEKYMMSL 61  
 260 GAVNPVTKFV--VNTDSSSVTNATSIQTAPASMLI-GDHYLCVYTA----- 306  
 62 LDRSQRLQIVLSPFLFVEDVMERORLIESVDPSTPLIYEETDWINIDIFH 121  
 307 ---TORISLQWL-----RRIGNY--SVMDICDYDESSGRW-NC----- 339  
 122 VPFQSHSEIEELFFASECKTGFRH-----LYKITSILKSKYKSSSGGLPAPSDKCP 174  
 340 LVARQHIETSTGWGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI-----CYFOID 390  
 175 IKEEIAITSGEVEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLRHLHYVSVNPGEVTRL 233  
 391 KKDCTITTKGTWVIG-----EALNSDYLYISNEKMGPGGRNLKYLQIDITYATCL 445  
 234 TDRGSHSCCIS-QHCDPFIISKYSNQKNPH--C---VSLYKLSPEDDPTCKTEFMA 285  
 446 -----SCLNPERCOYVSFSKEAKYQLRCSPGRLPYTLHSSVNDKGRVLED-N 497  
 286 TLIDSGRLPDYRPE---IFSFESTTGFTLYGMLYKPHDLOGKGYPTVLFYIGP--- 339  
 498 SALDKM--LQNVOMPSKGLDFIILNETKF-WYQMLPPH-FDKSKYPLLDVYAGPCSQ 553  
 340 ----- 339  
 Db 554 KADTVRLMAYTLASTENIIVASFQGRSGSGYGDKIMAHINRRLGTFVEVDIEAARQ 613

QY 340 -----QVAIAGAPYTLMIFYDTGYTERMG- 364  
 614 SKMGFVNDNRIRAIWGSYGYVTSMWLSGSGGVFKGIAPVSRREHYDSYTERMGL 673  
 Db 365 -HPDQEGGYLGSVAMQAEKPESEPNRLLLHGLDENVHAHTSILSLFLVRAKPYD 423  
 674 PFPELDLHRYNSTVMSRAENF--KQEVYLLHGTADNVHFGQSAQISKALVDGVDFQ 731  
 QY 424 LQIYPERHSIRVPESEHYELHLHYLOE 453  
 Db 732 AMWYTDHEDHIGASSTAHQHIYTHMSHFIRQ 761  
 RESULT 10  
 S23752  
 dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse  
 N/Alternate names: CD26 alpha subunit; THAM alpha subunit  
 C/Species: Mus musculus (house mouse)  
 C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 10-Sep-1999  
 R/Marguet, D.; Bernard, A.M.; Vivier, I.; Darmoul, D.; Naquet, P.; Pierres, M.  
 J. Biol. Chem. 267, 2200-2208, 1992  
 A:Title: CDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-c  
 A:Reference number: S23752; MUID:92129288; PMID:1370813  
 A:Accession: S23752  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-760 <MAR>  
 A:Cross-references: EMBL:X58384  
 R/Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;  
 J. Immunol. 147, 447-454, 1991  
 A:Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl) peptidase  
 A:Reference number: A46465; MUID:91302787; PMID:1112807  
 A:Accession: A46465  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 4-20 <WV>  
 A:Experimental source: M14.T thymoma cells, Swiss nu/nu  
 R/Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.  
 Biochemistry 33, 15204-15214, 1994  
 A:Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.  
 A:Reference number: A56030; MUID:95092780; PMID:799781  
 A:Accession: A56030  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 746-760 <BER>  
 A:Cross-references: GB:U12620  
 C:Genetics:  
 A:Gene: CD26  
 C/Superfamily: dipeptidyl-peptidase IV  
 C/Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein  
 F:213,223,315,514,679/binding site: carboxydrate (Asn) (covalent) #status predicted  
 F:624,702,734/Active site: Ser, Asp, His #status predicted  
 Query Match 9.3%; Score 230; DB 1; Length 760;  
 Best Local Similarity 22.6%; Pred. No. 3.6e-10;  
 Matches 98; Conservative 47; Mismatches 120; Indels 168; Gaps 19;  
 Db 140 KGFPHLYKITSLKSKYKSSSGGLPAPSDK-PIKEEIAITSGEVEVLGRHGSNIQV 198  
 372 KQGYKHIHF-----PKDKXDCFP-----ITKAMEVI-----SIBA 403  
 199 DEVRRLVYFEGT-KDSPLRHLHYVSVNPGEVTRLTDRGSHSCCIS---QHCDPFI 252  
 404 LNSDYLYISNQYKEMPGGRNLKYI-----QLTD--HTNVKCLSCDLNPERCOYTA 452  
 QY 253 SKYSNQNPH--C---VSLYKLSPEDDPTCKTEFMA TLIDSG---PLPDYRPE- 301  
 453 VSFSEKAKYQYQCGWPGRLPYTLHRSTDHKEALVLE-----DNSALDRMLQDVMPRK 506  
 QY 302 ---IFSFESTTGFTLYGMLYKPHDLOGKGYPTVLFYIGP----- 339

```

Db      507 KLDFFLYNTRF-WYQMLPPI-FDKSKYPLLDVYAPGSCGKADAPSLMWTATLAST 564
Qy      340 -----OVALAG-- 345
Db      565 ENIIASFDRGSGYOGDKIMAINRLGLEVEDOIEARQFVKMGFVDSKRVAIWGS 624
Qy      346 -----APVTLWIFYDTGTERYMG--HPDNEGGYIGSVAM 380
Db      625 YGGYVTSWLVGSGGVFKGCIAPAPSRMEYDSVTEYKGLPIPEDNLDHRYSTWS 684
Qy      381 QAEKPEPSEPNRLLLHGLDENVFAHTSILSLFLVRACKPYDLQIYPOERSIRVPSG 440
Db      685 RAEHF--KQVEYLHIGTADNVHFGQSAQISKVLVDAGVQAMMYTDEHGIASSTAH 742
Qy      441 EHYELHLHYLOE 453
Db      743 OHYSHMSHFLQO 755

```

## RESULT 11

```

T25173
hypothetical protein T23F1.7a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T25173
R:Wilkinson, J.
Submitted to the EMBL Data Library, October 1996
A:Reference number: Z19990
A:Accession: T25173
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-779 <MIL>
A:Cross-references: EMBL:Z81129; PDB: CAB03411.1; GSPDB: GN00023; CESP: T23F1.7a
A:Experimental source: clone T23F1
A:Gene: CESP: T23F1.7a
A:Map position: 5
A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1
C:Superfamily: dipeptidyl-peptidase IV

```

```

Query Match      9.1%; Score 226; DB 2; Length 779;
Best Local Similarity 21.1%; Pred. No. 7.7e-10;
Matches 99; Conservative 62; Mismatches 142; Indels 166; Gaps 18;

Qy      106 YEETDININIDIHVFPQSHHEIEFIIFASECTGRHLKYISILKSKYKSSGGL 165
Db      350 KYVASKRWVT-HDDPHSI-TSFEDTLFLLP-----HDKRDNAFOQVASLRSLSHGQL 399
Qy      166 PAPSDPKPIKEIATISGEMEVLRHGSNIQVDEVRRLVPEGTKDSPLBHLVYVSIV 225
Db      400 RTPK-----FLNIGEYVTSINGINKET---RTIFPHAAPKPSHSLFSYS-- 443
Qy      226 NPEVTRLTDRGYSHSCIS--QHCDFFISKYSNQKNPHCVSLYKLSPPD----- 274
Db      444 -----LADSRNAYCISCSIKNCITWAQOMDQMTAIVSCGPAAPHAIIVNLTRM 496
Qy      275 DPTCKTK-----EFMATILDSAGPLDPYTPPEIFSESTGTFTLGMVLYKPHDLQPEK 327
Db      497 DSDKTEHANLLYDXTYQNRVEAG-LPVYIKETI---KISDDFALIKLSIPKDIYNRD 552
Qy      328 KY--PTVLFYIGP----- 339
Db      553 KHQAIFLVHYVYGGPNDQNTKEATQIGIEVVASASQAAILRIDRGSGGSGMKYRSALY 612
Qy      340 -----QVALAG-- 345
Db      613 GOLGIVEVEDQIKAIKVLRLRYRHLLDARRAVAVFGMSYGGFWTLMSVNEAPQPFKCAVS 672
Qy      346 -APVTLWIFYDTGTERYMGHPDNEGGYIGSVAMQAEKPPSEPNRLLLHGLDENVH 404
Db      673 VAPVTNFAYVYATYTERYMG--DAPLESY--SDVTKKLDNFKS--TRLLMHGLLDNVH 726

```

```

Qy      405 FAHTSILSLFLVRACKPYDLQIYPOERSIRVPSGEGHVELHLHYLOE 453
Db      727 FQNSAILIDELQNRGVDPDLMYFPQASHLSRST--HYVGKMTHTFLRQ 773

```

## RESULT 12

```

T25174
hypothetical protein T23F1.7b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T25174
R:Wilkinson, J.
Submitted to the EMBL Data Library, October 1996
A:Reference number: Z19990
A:Accession: T25174
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-799 <MIL>
A:Cross-references: EMBL:Z81129; PDB: CAB03412.1; GSPDB: GN00023; CESP: T23F1.7b
A:Experimental source: clone T23F1
A:Gene: CESP: T23F1.7b
A:Map position: 5
A:Introns: 13/3; 52/1; 111/3; 197/3; 366/3; 524/1; 557/1; 648/2; 703/1; 743/1
C:Superfamily: dipeptidyl-peptidase IV

```

```

Query Match      9.1%; Score 226; DB 2; Length 799;
Best Local Similarity 21.1%; Pred. No. 8e-10;
Matches 99; Conservative 62; Mismatches 142; Indels 166; Gaps 18;

Qy      106 YEETDININIDIHVFPQSHHEIEFIIFASECTGRHLKYISILKSKYKSSGGL 165
Db      370 KYVASKRWVT-HDDPHSI-TSFEDTLFLLP-----HDKRDNAFOQVASLRSLSHGQL 419
Qy      166 PAPSDPKPIKEIATISGEMEVLRHGSNIQVDEVRRLVPEGTKDSPLBHLVYVSIV 225
Db      420 RTPK-----FLNIGEYVTSINGINKET---RTIFPHAAPKPSHSLFSYS-- 463
Qy      226 NPEVTRLTDRGYSHSCIS--QHCDFFISKYSNQKNPHCVSLYKLSPPD----- 274
Db      464 -----LADSRNAYCISCSIKNCITWAQOMDQMTAIVSCGPAAPHAIIVNLTRM 516
Qy      275 DPTCKTK-----EFMATILDSAGPLDPYTPPEIFSESTGTFTLGMVLYKPHDLQPEK 327
Db      517 DSDKTEHANLLYDXTYQNRVEAG-LPVYIKETI---KISDDFALIKLSIPKDIYNRD 572
Qy      328 KY--PTVLFYIGP----- 339
Db      573 KHQAIFLVHYVYGGPNDQNTKEATQIGIEVVASASQAAILRIDRGSGGSGMKYRSALY 632
Qy      340 -----QVALAG-- 345
Db      633 GOLGIVEVEDQIKAIKVLRLRYRHLLDARRAVAVFGMSYGGFWTLMSVNEAPQPFKCAVS 692
Qy      346 -APVTLWIFYDTGTERYMGHPDNEGGYIGSVAMQAEKPPSEPNRLLLHGLDENVH 404
Db      693 VAPVTNFAYVYATYTERYMG--DAPLESY--SDVTKKLDNFKS--TRLLMHGLLDNVH 746
Qy      405 FAHTSILSLFLVRACKPYDLQIYPOERSIRVPSGEGHVELHLHYLOE 453
Db      747 FQNSAILIDELQNRGVDPDLMYFPQASHLSRST--HYVGKMTHTFLRQ 793

```

## RESULT 13

```

A41793
dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A41793
R:Wada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimasaki, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A:Title: Differential expression of two distinct forms of mRNA encoding members of a di
A:Reference number: A41793; MUID:92108018; PMID:1729689

```

A:Accession: A41793  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-803 <WAD>  
 A:Cross-references: GB:M76429; NID:g408719; PIDN:AAQ41623.1; PID:g408720  
 A:Note: Sequence extracted from NCBI backbone (NCBIP:75138)  
 C:Superfamily: dipeptidyl-peptidase IV  
 C:Keywords: dipeptidylpeptide hydrolase, glycoprotein  
 F:257,342/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.9%; Score 221; DB 2; Length 803;  
 Best Local Similarity 20.4%; Pred. No. 2e-09;

Matches 107; Conservative 65; Mismatches 178; Indels 174; Gaps 19;

```

QY 44 YIARAGMTPEGKVA--WSILLDRSQRRLQIVLISPELFIPEDDVMERQRLIESVPSVY 101
DB 320 YIMVWMASTKVAVM--LSRAQN-----VS-----LITLCDAIT 353
QY 102 PLII--YEETTDIWINIHDFHVFPOSGHEIEFIFASECKTGFRHLKYKTSILK--ESK 157
DB 354 GVCCKHDESEAWL-----HRQNEPVPVS--KDG-RKFFVRAIPQGGQK 397
QY 158 YKSSGGLPAPSPFCPIKEIAITSGEVLGRHGSNIQDEVRLVYEGTKDSPLEH 217
DB 398 FHHITVSSQPNSSNDNIQ---SITSGMDVT---KILSYDEKSGQIFLSTEDLPRRR 450
QY 218 HLTVSVYVNGEVTR-----LTDGYSHCSCIQHCFEISKYSNQKPHCVSLYKLS 271
DB 451 QIYASASTV--GSFNRCGLCDLVNCTYFSASFSPGADFFLKCCEGPVP--TVSVNTTD 507
QY 272 PEDDPCKTKE-FWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYP 330
DB 508 KKKMFLETNEHVOKAISRQMPKVEYRKIE-----TDDYNLPIQILKPAFTTDTNAYP 561
QY 331 TVLFYIGCP----- 339
DB 562 LLLVVGCTGSSQVAEKFVATWETWVSSHGAVVVKCDGRSGSQGTLLHEVRRRLGL 621
QY 340 -----QVALAG-----APVT 349
DB 622 EEKQMEAVRMLKEPIIDIKTRVAVFGKDYGYLSTYLLPAKGQGAQAPVSSCGSALSPIT 681
QY 350 LMIFFYDTGTERMGHPDQNEQGYILGSVAMQAEKFPSEPNRLLLHGLFDENVHFAHTS 409
DB 682 DPELVASAFSERLGLHGLDNRAVEAKVAHRVSL--EQOGLVHTATADKIHQHTAELI 739
QY 410 ILSFLVRAGKPYDQIYPOERHSIRVPSGEHYELHLHYLQ 453
DB 740 ELITOLIKGKANYSLQIYDESHYFSSALQOHLHRSILQFVE 783

```

#### RESULT 14

168600  
 dipeptidyl aminopeptidase like protein - human  
 C:Species: Homo sapiens (man)  
 C>Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: I68600  
 R:Tokocant, N.; Doi, K.; Wenthold, R.J.; Wada, K.  
 Hum. Mol. Genet. 2, 1037-1039, 1993  
 A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat  
 A:Reference number: 154331; MUID:93372805; PMID:8103397  
 A:Accession: I68600  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-803 <RES>  
 A:Cross-references: GB:M96860; NID:g306707; PIDN:AAA35761.1; PID:g306708  
 C:Superfamily: dipeptidyl-peptidase IV

Query Match 8.7%; Score 216; DB 2; Length 803;  
 Best Local Similarity 19.1%; Pred. No. 5.1e-09;

Matches 111; Conservative 79; Mismatches 179; Indels 212; Gaps 23;

QY 2 GTANPKVTFKMSIMIDAGRIIDVIDKELIQFEILFEGVEYIARAGTPEGKVA--WS 59

```

DB 286 GSENPISILH-----VILNGPPTHDL---EMPPDDPRMREY-YITWVKWATSTKVAVTW- 336
QY 60 ILLDRSQTLQIVLISPELFIPEDDVMERQRLIESVPSVPLII--YEETTDIWINIH 117
DB 337 -LNRQN-----VS-----LITLCDAITGVCTKHEDESEAWL-- 369
QY 118 DIFHVPQSGHEBEI-----EFIFASECKTGR-HLYKTSILKESKYRSSGGLPAPSD 170
DB 370 -----RQNEPVPVSQDGRKFFIRAIPOGGRGKYHTT--VSSQPNSSNDNIQ----- 416
QY 171 FKCPIKEIAITSGEVLGRHGSNIQDEVRLVYEGTKDSPLEHLYVSVYVNGEV 230
DB 417 -----SITSGMDVT---KILAYDEKGNKIYFLSTEDLPRRRQLYSAN----- 456
QY 231 TLLTDGYSH--SCCISQHCDFEISKYSNQKPHCVSLYKLSR-----EDDPCKTK 281
DB 457 ---TEGNFRCGLCDLVNCTYFSASFHSD--FFLLKCEGPVPVMTVNTTDDKK 510
QY 282 EF-----WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVL 333
DB 511 MFDLETNEHVKKAINDRQMPKVEYRDI-----DDVNLPMQILKPAFTTDTNAYP 564
QY 334 FYIGCP----- 339
DB 565 VVDGTFGSSQVAEKFVSWETWVSSHGAVVVKCDGRSGSQGTLLHEVRRRLGLEEK 624
QY 340 -----QVALAG-----APVTLM 352
DB 625 DQMEAVRMLKQIYDRTKTRVAVFGKDYGYLSTYLLPAKENGQGTTCGALSPTIDPK 684
QY 353 FYDTGTERMGHPDQNEQGYILGSVAMQAEKFPSEPNRLLLHGLFDENVHFAHTSIL 412
DB 685 LVASAFSERLGLHGLDNRAVEMTVAHRVSL--EQOGLIHTPADKIHQHTAELI 742
QY 413 SFVLRAGKPYDQIYPOERHSIRVPSGEHYELHLHYLQ 453
DB 743 TOLIRGKANYSLQIYDESHYFSSSLKQHLYSIINFVE 783

```

#### RESULT 15

154331  
 dipeptidyl aminopeptidase like protein - human  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jul-2000  
 C:Accession: 154331  
 R:Yokotani, N.; Doi, K.; Wenthold, R.J.; Wada, K.  
 Hum. Mol. Genet. 2, 1037-1039, 1993  
 A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat  
 A:Reference number: 154331; MUID:93372805; PMID:8103397  
 A:Accession: 154331  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-865 <RES>  
 A:Cross-references: GB:M96859; NID:g306705; PIDN:AAA35760.1; PID:g306706  
 C:Superfamily: dipeptidyl-peptidase IV

Query Match 8.7%; Score 216; DB 2; Length 865;  
 Best Local Similarity 19.1%; Pred. No. 5.7e-09;

Matches 111; Conservative 79; Mismatches 179; Indels 212; Gaps 23;

```

QY 2 GTANPKVTFKMSIMIDAGRIIDVIDKELIQFEILFEGVEYIARAGTPEGKVA--WS 59
DB 348 GSENPISILH-----VILNGPPTHDL---EMPPDDPRMREY-YITWVKWATSTKVAVTW- 398
QY 60 ILLDRSQTLQIVLISPELFIPEDDVMERQRLIESVPSVPLII--YEETTDIWINIH 117
DB 399 -LNRQN-----VS-----LITLCDAITGVCTKHEDESEAWL-- 431
QY 118 DIFHVPQSGHEBEI-----EFIFASECKTGR-HLYKTSILKESKYRSSGGLPAPSD 170
DB 432 -----RQNEPVPVSQDGRKFFIRAIPOGGRGKYHTT--VSSQPNSSNDNIQ----- 478

```

```

QY 171 KCPKFEKAIASGEVLEVGRGSGNIQVDEVRVLYVEFGKSPLEHNLVYVYVNGEV 230
Db 479 -----SITSGMDVTV-----KILAYBEKGNKIFLSTEDIPRRRLYSAN----- 518
QY 231 TRLTRDGRYSH---SCCISQCHDFISKYSNOKNPHCVSLYKLSNP-----EDDPCTKT 281
Db 519 ---TEGNFNROCTSLDVLNCTYFASFSFHSMD---FLLKCEGEPVMTVYANTITDDKK 572
QY 282 EF-----WATIIISAGPLPDYTPPELFFSEFSTTGFLLYGMLYKPHDIQPKKYPYVL 333
Db 573 MFDELTEHNEHVKKAIINDROMPKVEYRDIET-----DDYNLPMQILKPATFTDTHYPLLL 626
QY 334 FIYGGP----- 339
Db 627 VVDGTPGSOSVAKEFEVSWETVMWSHGAVVVKCDGRSGFOCTKLHIEVRRRLGLIEEK 666
QY 340 -----QVAIAG-----APYTLMI 352
Db 687 DQMEAVRMTLKEQYIDRTAVAVAFGKDYGYLSTYILPAKGNGOQRTFTGSCASPIITDFK 746
QY 353 FYDTCYTERYMGHPQONEGGYLYGSAVMAQEKFPSPBNRLLLHGLDENVAHFAITSILL 412
Db 747 LYASAFSEERYILGHDGNRAVEMTKVAHVSVSL--EEQOFLIIHPADREKIHFOHTABELI 804
QY 413 SFLVRAKRPYDLOIYPOERHSIRVPESGHEYLHLHLIYLOE 453
Db 805 TOLIRKANYSLOIYPDSESHYFTSSLKOKHLYRSIINPVE 845

```

Search completed: October 15, 2003, 17:14:50  
Job time : 16.2186 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 (Search time 7.37729 seconds  
(without alignments)  
2964.152 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482

Sequence: 1 TGTANPKVTFKMSIWDAB.....HLHTYQENLGSRIALAKVI 465

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	11.2	818	1 DAP2_YEAST	P18962 saccharomyc
2	267.5	10.8	931	1 ST13_YEAST	P33894 saccharomyc
3	267	10.8	793	1 YEAB_SCHPO	O14073 schizosacch
4	262	10.6	765	1 DPP4_BOVIN	P81425 bos taurus
5	257	10.4	765	1 DPP4_FELCA	O92317 felis silve
6	249.5	10.1	767	1 DPP4_RAT	P14740 ratius norv
7	224	9.4	766	1 DPP4_HUMAN	P27487 homo sapien
8	221	8.9	760	1 DPP4_MOUSE	P28843 mus musculu
9	221	8.8	804	1 DPP4_BOVIN	P42659 bos taurus
10	218	8.8	804	1 DPP4_MOUSE	O92318 mus musculu
11	216.5	8.7	761	1 SEPR_MOUSE	P97321 mus musculu
12	216	8.7	865	1 DPP4_HUMAN	P42658 homo sapien
13	214	8.6	760	1 SEPR_HUMAN	O12884 homo sapien
14	214	8.6	859	1 DPP4_RAT	P46101 ratius norv
15	145.5	5.9	853	1 YDZF_SCHPO	O92799 schizosacch
16	111.5	4.5	1019	1 CAZ6_HUMAN	P12110 homo sapien
17	104	4.2	1436	1 DPP3_STRAH	O53665 staphylococ
18	104	4.2	1438	1 DPP3_STRAH	O95048 staphylococ
19	103.5	4.2	1435	1 DURI_YEAST	P32328 saccharomyc
20	103.5	4.2	4451	1 GR8B_BACBR	P14688 b gramicidi
21	102	4.1	732	1 ACER_MOUSE	P22667 mus musculu
22	102	4.1	1312	1 ACER_MOUSE	P09470 mus musculu
23	101	4.1	1313	1 ACER_RAT	P47820 ratius norv
24	100	4.0	558	1 LCBI_YEAST	P25045 saccharomyc
25	99.5	4.0	543	1 PAT2_YEAST	P38137 saccharomyc
26	97.5	3.9	984	1 DPOL_NPVAC	P18131 autocographa
27	96.5	3.9	2560	1 PPS2_BACSU	P39846 bacillus su
28	96	3.9	413	1 CSD_THEMA	O94191 thermocoga
29	96	3.9	581	1 PRLR_SHEEP	O46661 ovis aries
30	95.5	3.8	1029	1 CAZ6_MOUSE	O02788 mus musculu
31	95	3.8	692	1 Y650_METUA	O57852 methanococc
32	94.5	3.8	570	1 NCAP_TACV	P18140 tacaribe vi
33	94	3.8	550	1 INVI_HANAN	P40912 hansenula a

34	94	3.8	1014	1 BGAL_BACHD	O9K9C6 bacillus ha
35	93.5	3.8	716	1 YDQA_SCHPO	O13730 schizosacch
36	93.5	3.8	1203	1 SDCL_CABEL	P24349 caenorhabdi
37	93	3.7	732	1 ACPI_HUMAN	P13798 homo sapien
38	93	3.7	998	1 RRPD_FHV	O66929 flook house
39	93	3.7	1233	1 VLI_REOVD	P15024 reovirus (c
40	92.5	3.7	506	1 GAT4_TREPA	O83983 treponema p
41	91.5	3.7	761	1 YKQ2_CABEL	P41993 caenorhabdi
42	91.5	3.7	1461	1 NEOL_HUMAN	O92859 homo sapien
43	91	3.6	986	1 DPOL_NPVEM	P41712 bombyx mori
44	90.5	3.6	375	1 ADH_GADCA	P26325 gadus calla
45	90.5	3.6	2212	1 RRP1_EBOZM	O05318 ebola virus

## ALIGNMENTS

RESULT 1  
DAP2\_YEAST STANDARD; PRT; 818 AA.  
ID DAP2\_YEAST  
AC P18962;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE dipeptidyl aminopeptidase B (EC 3.4.14.-) (DAP B) (YSCV).  
GN DAP2 OR YHR028C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_Taxid=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89174971; PubMed=2647766;  
RA Roberts C.J., Pohlig G., Roehman J.H., Stevens T.H.;  
RT "Structure, biosynthesis, and localization of dipeptidyl  
RT aminopeptidase B, an integral membrane glycoprotein of the yeast  
RT vacuole.";  
RL J. Cell Biol. 108:1363-1373(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
Lacaille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,  
Nhan M., Rifkin L., Riles L., St Peter H., Trevasaki E., Vaughan K.,  
Vaughan M., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
RT VIII.";  
RL Science 265:2077-2082(1994).  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE  
CC VACUOLES.  
CC -1- SIMILARITY: Belongs to peptidase family S9B.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X15484; CAA33512.1; -  
CC EMBL: U10399; AAB68879.1; -  
CC PIR: S46780; A30107.  
CC  
CC MEROPS: S09.006;  
CC SGD: S0001070; DAP2.  
CC GO: GO:000329; C:vacuolar membrane (sensu Fungi); IDA.  
CC InterPro: IPR002469; DDP1V N term.  
CC InterPro: IPR001375; Peptidase S9.  
CC InterPro: IPR002471; Prol\_endopep\_ser.

```

DR InterPro: IPR000379; Ser. esterase site.
DR Pfam: PF00930; DppIV N-term: 1.
DR Pfam: PF009326; Peptidase S9: 1.
DR PROSITE: PS00708; PRO-ENDOPEP-SER; 1.
DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 29
FT TRANSMEM 30 45
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT LIPIDAL (POTENTIAL).
FT ACT_SITE 46 818
FT ACT_SITE 679 679
FT ACT_SITE 756 756
FT ACT_SITE 789 789
FT CARBOHYD 63 63
FT CARBOHYD 79 79
FT CARBOHYD 110 110
FT CARBOHYD 139 139
FT CARBOHYD 372 372
FT CARBOHYD 392 392
FT CARBOHYD 421 421
FT CARBOHYD 738 738
FT CONFLICT 83 83
FT CONFLICT 125 125
FT CONFLICT 182 188
FT CONFLICT 200 200
FT CONFLICT 366 375
FT CONFLICT 808 818
SQ SEQUENCE 818 AA; 93404 MW; 318F45045375BD3 CRC64; .

Query Match 11.2%; Score 277; DB 1; Length 818;
Best Local Similarity 21.0%; Pred. No. 2.2e-14;
Matches 121; Conservative 64; Mismatches 170; Indels 220; Gaps 22;

48 ACHTEPGKXAMSLRLRSQRLQIVLSPFLFVDDWDERRLIESVDSTPLIYE 107
251 AMWSPGDIYALFKIDESEVGERIT---PYVDEKDIYENRSLK-YKSGTP---N 301
108 ETTDIWI-NIHD--IHFVFPQSHEEIEFIF-----A-SECKTGFRHLKYIT 150
302 PHAELWVGMKDKTSFHRISGNKKKDSLLITEVTWNGNVLVKTTRSSDILTYFLID 361
151 SLIKESKYR---SSGG-----LPAPSPFKPIKEI----- 179
362 TIAKTNVNVNNESSNGMWEITNTLFIPIANETFDPRHNGYVILPIGYNHLAYENS 421
180 -----AITSGEVNLGRHSGNIQVDEVRRLVFEFGTKDPSLEHLVVSVNPGVTRLT 234
422 SSHYKTLTEGKMEV---NGPLAFDSMENRFLFISTRKSTENHVIYIDLRSENEIEVT 478
235 DRGYS--HSCCISQHCDFPISKYSNOKNPH-----C-----VSLYKLSGP 272
479 DTSEGDVYDVFSFGSRFGLLTAKGPKVFOYKIVDFHSRAKAECDKGNVLGKSLYLUEN 538
273 EDDPTCKTEFMAITIIDASGRLDYTPPELFSESTTGFLYLG--MLYKPHDIQ----- 325
539 E-----VLTGI--LEDYAVPRKSRFRELMLGKDFEKOILVMSYEILPNDPDE 583
326 --GKKYPTVLFIYGP----- 339
584 TISDHFVFFFAVAGNSQGVVTFSGVGFNEVVAQLNATLVVVVDGRGTGFKGDRSLV 643
340 -----QVALAG----- 345
644 RDRLDGVEARDQISASLSYGLTFVDPQKISLFGWSYGYLTLLKTEKQGRHFKYGMV 703
346 APVTLMIPTDGTGTEYMGHPDNEGYIGSV-----AMQAEKFPSEPRNLILHGFIDE 401
704 APVTLDRFYDSVYTERYMHTPQSNFPGYVSSVHNVATL-----AQAANFLMHGTGDD 757
402 NVHFASTSLSLFLVRAG--KPYDLQIYPCERHSIR 435
758 NVHFNLSLKLFLDLDDLNGVENVYDVHVFPPSDHSIR 792

```

```

RESULT 2
ST13_YEAST
ID ST13_YEAST STANDARD; PRT; 931 AA.
AC P33894;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DPA A) (YSCIV).
GN STE13 OR YC11 OR YOR219C OR YOR50-9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9506382; PubMed=7975897;
RA Anna-Artiola S.S., Herskowitz I.;
RT "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
RT aminopeptidase."
RL Yeast 10:801-810(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=AB320;
RA Flanagan C.A., Thorner J.;
RT "STE13."
RL (in) Getting M.-J., Novick P., Stevens T.H., Rochebiact J. (eds.);
RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL Press, Oxford (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=8286C / FY1679;
RX MEDLINE=9643797; PubMed=8840505;
RA Galsen F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae."
RL Yeast 12:877-885(1996).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC ALPHA-FACTOR PRECURSOR.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC VACUOLUS.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC EMBL: L21944; AAA5119.1; -
CC EMBL: U08230; AAA17897.1; -
CC EMBL: X52441; CAA63182.1; -
CC EMBL: 275127; CAA99437.1; -
CC PIR: A49737; A49737.
CC MEROPS: S09.005; -.
CC SGD: S0005745; STE13.
CC GO: GO:0005802; C:Golgi trans-face; IDA.
CC GO: GO:0041777; P:aminopeptidase activity; IDA.
CC GO: GO:007323; P:peptide pheromone maturation; IDA.
CC InterPro: IPR002469; DppIV N-term.
CC InterPro: IPR001375; Peptidase S9.
CC InterPro: IPR002471; Prol endopep ser.
CC InterPro: IPR000379; Ser-esterase site.
CC Pfam: PF00930; DppIV N-term: 1.
CC Pfam: PF009326; Peptidase S9; 1.
CC PROSITE: PS00708; PRO-ENDOPEP-SER; FALSE-NEG.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT DOMAIN 1 119
FT TRANSMEM 120 140
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

```

```

FT DOMAIN 141 931 LUMENAL (POTENTIAL). (BY SIMILARITY).
FT ACT SITE 785 785 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 863 863 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 896 896 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 931 AA; 107200 MW; 81AF70094093C023 CRC64;

Query Match 10.8%; Score 267.5; DB 1; Length 931;
Best Local Similarity 21.3%; Pred. No. 1.5e-13;
Matches 114; Conservative 63; Mismatches 164; Indels 195; Gaps 18;

OY 74 ISPELFIPEVDVEMERORLIESVP-----DSVPLLIYEETTDI---WI-NHIDIFHVFQ 125
DB 445 ISPDFE---RFEITRNSKILDVKYDIPSSQMLVTRNTNSLFGMWIEKTDILSIPKX 501
OY 126 SHEEIE-FIPASCKTGFRHLKYITSLKSKYKRSGLPADSDFKPIKEIATIS 183
DB 502 PELKMDYGDYIDHADSRSFSLFYFVYF-----AKEPIQLTK 540
OY 184 GEMEVLGRH--GSNIQVDEVRRLVFEETKDSPLNEHLVYVSV-----NPG 229
DB 541 GMEVLTGNGIYGEYETD---TTFETANEIGVMSQHLYSISLDTSTQNTFQSLQNP-- 594
OY 230 VTRLRGVSHSCCISCHDFFISKYSNKNP-----HCVSLYKLSPE 273
DB 595 ---SDKXDFYDFELSSARAYISKKGPDFTIKVAGPLTRVLANAEIHDSILQITDE 650
OY 274 DDPTKTKTFEATILDSAGPLPDYTPPELFESESTT---GFLYGLMYKPHDLOPKKYP 330
DB 651 ---KFE-----KINYDLP-ITSYKTVLDDGVAINIYIEIKPAILNPKKYP 694
OY 331 TVLFIYGP----- 339
DB 695 ILVNIYGGPSOTFTTKSGLAEQAVSGLDVIVLQIEPRGTGKGMSFRSAREKLGW 754
OY 340 -----OVAIAGAPYTL 350
DB 755 EPRDITEVTKKFIQNSQIHDESKIAIMQWSYGFSLKTVELDNGDFPKYMAVAAPTN 814
OY 351 WIFDPTGTERYMGPPDQNEOGYIYGVAMQAEKPPSEBNLLHGLFIDENVAHRTSI 410
DB 815 WTLVDVSYTERYMNQPSNHEGCFEVSTIQNKSEES-LKRLFIYHGFDDVHVIQNTFR 873
OY 411 LLSFLVRAG-KPYDQIYPOERSIRVPSGEHYELHLHYQENLGSRIALAKYI 465
DB 874 LVQDLNLGLINVDHMFIPDSDHSIRYNAQRIYFQKLYWLRLDAERFNTETVL 929

RESULT 3
YEAR SCHPO STANDARD; PRT; 793 AA.
ID YEAR SCHPO
AC 014073;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
GN SPACUNK4.08 OR SPAC2E11.08 (EC 3.4.14.-).
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI TaxID=4896;
RN NCB1_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; Pubmed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hilsago J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.U., Hunt S., Jagers K.,

```

```

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmons M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Meestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purrelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpukovski G.V., Useery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC - SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC - SIMILARITY: Belongs to peptidase family S9B.
CC - SIMILARITY: Belongs to peptidase family S9B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, AL031180; CAA20138.1; -.
CC PIR, T41703; T41703.
CC GenBank, SPombe; SPACUNK4.08; -.
CC InterPro; IPR002469; DPEIV_N term.
CC InterPro; IPR001375; Peptidase S9.
CC InterPro; IPR002471; Prol endopep ser.
CC InterPro; IPR00379; Ser_cyste_site.
CC Pfam; PF00930; DPEIV_N term; 1.
CC Pfam; PF00326; Peptidase S9; 1.
CC DR Pfam; PF00326; Peptidase S9; 1.
CC DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
CC KW Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
CC Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
CC CYTOPLASMIC (POTENTIAL).
CC DOMAIN 1 24
CC TRANSMEM 25 45
CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC LUMENAL (POTENTIAL).
CC CHARGE RELAY SYSTEM (BY SIMILARITY).
CC CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT SITE 755 755 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT SITE 755 755 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 793 AA; 91304 MW; 20B70FP97F231463 CRC64;

Query Match 10.8%; Score 267; DB 1; Length 793;
Best Local Similarity 25.6%; Pred. No. 1.4e-13;
Matches 100; Conservative 37; Mismatches 127; Indels 126; Gaps 14;

OY 179 IATSGEMEVLGRGNSIQVDEVRRLVFEETKDSPLNEHLVYVSYPNGEVTLRDGY 238
DB 412 IYLTSGAMDYV---DGPRIHIDGDFGNVYFLATLKDSTERHLVYVS-LDTLEIYGTIDNGE 467

```



QY	229 S----	HSCCISGCDFFISIXYXNOKNHCVCISLYLSSPEDDPCTKTEFPAFTLIDSAGPLR	295
Db	468	DEGYISTSPGCDFFVLYNHPDV-----WQELRSTKDXDYCLSLFTNSRLKQO---LS	520
QY	236	DYTPPEI-----FSFESTTGFTLYGLMYKRPDLQPGKKYPTVLFYSGP-----QYA	342
Db	521	SITLPSVEYGLKTFNPTT-FNF--MERRBRNFDVKKYVPVLPFAYGCGPSQVAKLFRVD	577
QY	343	-----	342
Db	578	FOAYLASHDFEPIVTLTDRGTGENGAFRYSVSRHLGEMESYDQAGAKFMADLPFVD	637
QY	343	-----IAGAPVTLWIFYDGTGYTERYMGHPDQNEQY	373
Db	638	ENHVGIWGSYGGYLTLKLTLETQDVPFSYGMAVAPVTDWRLYDSVTERVYVDLPQYKKEG-	696
QY	374	YLGSAVMAQKAEKPPSEPNRLLHLGFLDENVHPHSTILLSFLVRAG-KPYDLOIYQERH	432
Db	697	YNSQSIHDYEKE-KOLKRFVVAHGTDDBDVHFFQHSNMLMDGLNLANCYNNDMAVFPDSAH	755
QY	433	SIRVESEGHYEHLHLHYQENLGSRIAL	462
Db	756	SI-----SYHNASLSTYHRLSEWIGDALGRI	781
RESULT 4			
DP4	BOVIN		
ID	DP4_BOVIN	STANDARD;	PRT; 765 AA.
AC	P81425; O8BMG8;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Dipectidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation		
DE	antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)		
DE	(Activation molecule 3) (ACT3) (WC10).		
GN	DP4 OR CD26		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;		
OC	Bovidae; Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Lymphocytes;		
RX	MEDLINE=22067734; PubMed=12073152;		
RA	Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,		
RA	Bohach G.A.;		
RT	"Molecular characterization of bovine CD26 upregulated by a		
RT	staphylococcal superantigen.";		
RL	Immunogenetics 54:216-220(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.		
RC	TISSUE=Thymus;		
RX	MEDLINE=22021197; PubMed=11981836;		
RA	Gilddon D.R., Howard C.J.;		
RT	"CD26 is expressed on a restricted subpopulation of dendritic cells in		
RT	vivo.";		
RL	Eur. J. Immunol. 32:1472-1481(2002).		
RN	[3]		
RP	SEQUENCE OF 1-24.		
RC	TISSUE=T-cell;		
RX	MEDLINE=21482004; PubMed=11598101;		
RA	Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,		
RA	Naessens J., Bohach G.A.;		
RT	"Density of activation molecule 3 on superantigen-stimulated bovine		
RT	cells is CD26.";		
RL	Infect. Immun. 69:7190-7193(2001).		
RN	[4]		
RP	SEQUENCE OF 537-546.		
RC	TISSUE=Kidney;		
RX	MEDLINE=98293306; PubMed=9629661;		
RA	Ben-Shooshan I., Parola A.H.;		
RT	"The Cp-I subunit of adenosine deaminase complexing protein from calf		

```

RT   kidney is identical to human, mouse, and rat dipeptidyl peptidase
RT   id.";
RL   Comp. Biochem. Physiol. 119B:289-292(1998).
CC   -I- FUNCTION: Removes N-terminal dipeptides sequentially from
CC   polypeptides having un substituted N-termini provided that the
CC   penultimate residue is proline. Binds and regulates the activity
CC   of ADA.
CC   -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC   Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC   Xcc is neither Pro nor hydroxyproline.
CC   -I- SUBUNIT: Homodimer.
CC   -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC   a soluble form (By similarity).
CC   -I- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
CC   several immune system tissues.
CC   -I- PTM: The soluble form (SDP) derives from the membrane form (MDPP)
CC   by proteolytic processing (By similarity).
CC   -I- SIMILARITY: Belongs to peptidase family S9B.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; AF461806; AAL67836.1;-
DR   EMBL; AY056834; AAL23628.1;-
DR   MEMOPS; S09.003;-
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR003079; Ser_ester_site.
DR   Pfam; PF00930; DPPIV_N term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1
KM   Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
KT   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN          1               765
FT                  DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                  (MDPP).
FT   FT             38              765
FT   DOMAIN         1               6
FT   TRANSMEM      7               29
FT                  SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                  (POTENTIAL).
FT   DOMAIN         30              765
FT   ACT_SITE      629              629
FT   ACT_SITE      707              707
FT   ACT_SITE      739              739
FT   CARBOHYD      84               84
FT   CARBOHYD      91               91
FT   CARBOHYD     149              149
FT   CARBOHYD     218              218
FT   CARBOHYD     228              228
FT   CARBOHYD     271              271
FT   CARBOHYD     280              280
FT   CARBOHYD     320              320
FT   CARBOHYD     392              392
FT   CARBOHYD     495              495
FT   CARBOHYD     684              684
SQ   SEQUENCE       765 AA;  88369 MW;  E31265421F43E116 CRC64;
Query March           10.6%; Score 262; DB 1; Length 765;
Best Local Similarity 21.7%; Pred. No. 3.3e-13;
Matches 124; Conservative 63; Mismatches 196; Indels 188; Gaps 20;
Qy    2 GTANPKYTFKMSLEIMDAEGRIIDVIDKELIQPELLIEFGVEVYIARAQWTPGCKYAAWSIL 61
Db    259 GAAMP--TIKFVNYSLSLSPNINATSOQLIPVPSALI-GDHVICDTWTVEERIS----- 311
Qy    69 LDRSQTLLQVLVLSPELFIPVEDDMRQRILSEVPSSVPLTLIIYEETTDWINIHDIHF 121
Db    312 -----LOMLRRIOYV--SIMDICDYDSGTWRMISSVGROH 344

```

```

OY 122 VEPQSHHEE-----EFIPASECKTGFRHLKYTSLKESKYKRSGLPAPS 169
DB 345 I-----EISTGWGFRPAPRPHFTSNGS-----FYILIS--NEBGKHI-----C 384
OY 170 DFKCPKIEIAITSGEWEVLGRHSGNIQVDEVARLVYFEGT-KDSELEHLLVYVSVNG 228
DB 385 HFQTDKRNCTFITKGAMEVIG-----IEALTSIDYLYISNEYKMGGANLKYIQLNDYTX 439
OY 229 EYTRLT-----DRGSHSCCISQHCDFPISKYSNQNPHCVSLYKLSSEDDPTCTKTFP 283
DB 440 KATCLSCELNPRCOYVSFSQEARVYQLRCSGRLP-----LYTLHNSNNKDELAVLE- 494
OY 284 WATILDSAGPLPD-YTPPEIFSESTGTFTL-YGMLYKPHDLQPKKYVTLFTYIGP- 339
DB 495 --NNSDLDOVLQVQMPKSKLDPIHLHGKTFWQMLPRPH-FDKSKKYTLLELVYAGPFS 551
OY 340 ----- 339
DB 552 OKADALFRNLWATYLASTENIIVASPDGRSGYGQDKIMHAINRRLGTFEVEDQIEATRO 611
OY 340 -----QVALAGAPVTLMIFYDGTTERVNG 364
DB 612 FSKMGFVDDKRIAIMGWSYGYVTSMVLGAGGVFKGIAVAPVSKMEYVDSVYTERVNG 671
OY 365 --HPDNEGGYLYGSVAMQAEKFPSEPNRLLLHGFLENVHFAHTSILSLFVRAKGPY 422
DB 672 LPTPEBNLDSYNSSTWSPRAENF--KQVEYLHIGRADNVHFOQSAQISKALVDAGVDF 729
OY 423 DLQIYPOERHSIRVPSGGEHYELHLLYLOE 453
DB 730 QSMWYTDDEDHGIASSFAHQHYTHMSHFLKQ 760

RESULT 5
DBP4_FELICA STANDARD; PRT; 765 AA.
ID DP4_FELICA
AC 09N217;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipterydyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
  antigen CD26).
DN DPP4 OR CD26.
OS Fells silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=peripheral blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
  Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
  activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368(1999).
CC -!- FUNCTION: Removes N-terminal dipeptides sequentially from
  polypeptides having unsubstituted N-termini provided that the
  penultimate residue is proline (By similarity).
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|
  Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
  Xcc is neither Pro nor hydroxyproline.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
  a soluble form (By similarity).
CC -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
  by proteolytic processing (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.iesb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB023952; BAA92344.1; -.
DR MEROPS; S09.003; -.
DR InterPro; IPR002469; DPPIV N term.
DR InterPro; IPR001375; Peptidase S9.
DR InterPro; IPR002471; Pro_endopep_ser.
DR InterPro; IPR000379; Ser_estr_site.
DR Pfam; PF00930; DPPIV N term; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
  Transmembrane; Glycoprotein; Signal-anchor.
  DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
  CHAIN 1
  FT 38 765
  FT 1
  FT DOMAIN 1 6
  FT TRANSMEM 7 29
  FT 30 765
  FT ACT SITE 629 629
  FT ACT SITE 707 707
  FT ACT SITE 739 739
  FT CARBOHYD 84 84
  FT CARBOHYD 91 91
  FT CARBOHYD 149 149
  FT CARBOHYD 178 178
  FT CARBOHYD 228 228
  FT CARBOHYD 280 280
  FT CARBOHYD 320 320
  FT CARBOHYD 330 330
  FT CARBOHYD 331 331
  FT CARBOHYD 519 519
  FT CARBOHYD 684 684
  SQ SEQUENCE 765 AA; 88213 MW; 3EFC96A22B175D9 CRC64;

Query Match 10.4%; Score 257; DB 1; Length 765;
Best Local Similarity 21.8%; Pred. No. 8,3e-13;
Matches 127; Conservative 63; Mismatches 182; Indels 210; Gaps 25;

OY 2 GTANPKVTFKMEIMDAEGRITIDVLDKELQPELFFGVE-IADAGTPEKVMSTL 61
DB 259 GAANP--TVKLVITDNLNPTNATSVETPPAAL-TGDYYLCQVTANERIS---- 311
OY 62 LDRSQTRLQIVLISPELFIPEVDWNEROLIESVDSVTPLIYETTDIMINIDIFH 121
DB 312 -----LQWLRLQNY--SYMDTRDYNSTGKMS----- 338
OY 122 VEPQSHHEEIEFIPASECKTG-----RHLKYTSLKESKYKRSGLP 166
DB 339 ---SAAQEHIEW-----STTGWVGRFRPAPRPHFTSDGRFXYKILS--NEDGYKHI----- 383
OY 167 ASDPKPKIEIAITSGEWEVLGRHSGNIQVDEVARLVYFEGT-KDSELEHLLVYVSV 225
DB 384 --CRFQIDKDKDCTFITKGAMEVIG-----IEALTYDLYISNEYKMGGRGLYKI--- 433
OY 226 NFGEVTRLDRGYSSCCIS-----QHCDFPISKYSNQNPH---C-----VSLYKLSPE 273
DB 434 -----QLND--YTVACLSCELPKRCOYVSFSKAKYQLRCSGGGLPYTLHRS 485
OY 274 DDPTKTFEWMATILDSAGPLDYTPPE---IFSSTGTFTLYGMLYKPHDLQPKKYP 330
DB 486 NDEERVLVED-NSALDKM--LQEVQMPKSKLDPIHLNTEKF-WYQWILPRPH-FDTSKKYP 540
OY 331 TYLFTYIGP----- 339
DB 541 LILIDVYAGCSQKADALFRNLWATYLASTENIIVASPDGRSGYGQDKIMHAINRRLGT 600
OY 340 -----QVALAGAPVTLMIFY 353

```

```

Db      601 EVEQOIEAARQFSKMGFPDDKRIAIIMWGSYGGYVTSMTVLGASGVFKGIAVAPYSRMEY 660
Oy      354 YDTCGTESEYMGHP-DQNECGYVLLGSVVA-QAEKPFSEENRLLHLGFIDENVAHPTSIL 411
Db      661 YDSYTEETTEYMWGLPPQDNLDDYVKNSTVSRANRF-KOVEYLLHGTADNVHFOQSQOI 718
Oy      412 LSFVFRAGKPYDLQIYPOERSHSIRPESGEHYELHLHYLOE 453
Db      719 SKALVDAGVDVFOAMWYTEDHGIASGPAHQHLYTMSHFHQ 760

RESULT 6
DPP4_RAT ID _DPP4_RAT STANDARD; PRT; 767 AA.
AC P14740:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (Gp110 glycoprotein) (Bile canaliculus domain-specific
DE membrane glycoprotein).
DE DPP4 OR CD26.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA RX MEDLINE=89123496; PubMed=2563382;
RA Ogata S., Misumi Y., Ikehara Y.,
RT "Primary structure of rat liver dipeptidyl peptidase IV deduced from
RT its cDNA and identification of the NH2-terminal signal sequence as
RT the membrane-anchoring domain."
RL J. Biol. Chem. 264:3596-3601(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA RX MEDLINE=88068516; PubMed=3479775;
RA Hong W., Doyle D.;
RT "cDNA cloning for a bile canaliculus domain-specific membrane
RT glycoprotein for rat hepatocytes."
RL Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
RN [3]
RP SIGNAL-ANCHOR
RA RX MEDLINE=90338089; PubMed=1974258;
RA Hong W., Doyle D.;
RT "Molecular dissection of the NH2-terminal signal/anchor sequence of
RT rat dipeptidyl peptidase IV."
RL J. Cell Biol. 111:323-328(1990).
RN [4]
RP SEQUENCE OF 281-302.
RC TISSUE=Kidney;
RX MEDLINE=94128239; PubMed=7905271;
RA Iwaki-Egawa S., Watanabe Y., Fujimoto Y.;
RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT dipeptidyl peptidase IV."
RL Biol. Chem. Hoppe-Seyler 374:973-975(1993).
RN [5]
RP FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
POLYPEPTIDE RESIDUE IS PROLINE.
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
A SOLUBLE FORM.
CC -1- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
BY PROTEOLYTIC PROCESSING.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

Query Match	Score	249.5	DB 1	Length	767
Basic Local Similarity	20.9%	Pred. No. 13-12			
Matches 124	Conservative 67	Mismatches 174	Indels 227	Gaps 26	
Qy 2	GTANPKTTPKMSLEIMDAEGRIDVIDKEILOPELIFEGVEVYIARAGWPECKYASML	61			
Dp 258	GAANPPTKFFI--VNTDSLSSTTTIMQITAPASV-TGDHYLCDAVANSSEDIIS----	310			
Qy 62	LDRSQTRLOIVLSPFLFIVEDDWERQRLIESVDPSTPLIYEFTTDMINIHDFH	121			
Dp 311	-----LQMLRRIQNY--SYVAICDYDKTTLVW-NC-----	337			
Qy 122	VFPQSHHEEIEFIPASEC-----KTGFRLLYKITSILSKSKY	158			
Dp 338	---PTQHEHTSTKTCGCFRPAEHPFTSDGSSFYIYVDDKQCYK-ICQYQKQK-	390			
Qy 159	KRSSGGLPAPSDPKPIKEEIA--ITSGEWEVLGRHNSNIQVDEVRLLVYFEET-KDSP	215			
Dp 391	-----PEQVCTFITIGAAEVL-----SIEALTSIDVLYIISNEYKEMP	428			
Qy 216	EHLLYVVSYNPPEVTRILTRGSHSGCIS-----QHCDPF--ISKYSQKPHC-----	263			
Dp 429	GRNLYKI-----QLTQ--HTNKKCLSCDINPEKQYIYVSLSEAYYUOLGCRGPG	477			
Qy 264	VSILYKSSPEDDPTCKTKEFWATILDSAGFLPDYTPPE--IFSFESTTGFTLYGLMYK	320			

Db 478 LPLYLHRSDDQKEHRLVED--NSALDKM--LQDVMPKSLDFIVLNTRF-WYQWILPP 533  
 Qy 321 HDLQFGKXPTVLFYGGP----- 339  
 Db 534 H-FDKSKYPLLDIVAGCSQKADAFRLNATYLASTENIIVASFDGRSGYGGDKIM 592  
 Qy 340 -----OVAIAG----- 345  
 Db 593 HAINRGLTLEVEDQLEAARQFLKMGFVDSKXVAIWMGSGYVTSMTLSSGSGVFKCGI 652  
 Qy 346 --APVTLMIFVDTGTYTERYMG--HPDQNEQGYLGSVAMQAEKFPSEPNRLILHGFIDE 401  
 Db 653 AVAPYSRMEYDYSVYTERYMGLPTEPDNLHDYRNSYVMSRAENF--KQVEYLIIHGTDAD 710  
 Qy 402 NVHFPHSTILSLFLVRACKPYDLQIYPOERHSIRPSEGEHEHLHLHYLQE 453  
 Db 711 NVHFQOASQISKALVDAGVDFOAMWYTTDEDHGLASSTAHQHYHSMSHFLQO 762  
 RESULT 7  
 DPP4\_HUMAN STANDARD; PRT; 766 AA.  
 AC P27487;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (TSP103) (Adenosine deaminase complexing protein-2) (ADABP).  
 GN DPP4 OR ADCP2 OR CD26.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=92329551; PubMed=1352704;  
 RA Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.;  
 RT "Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine proteinase on the cell surface.";  
 RL Biochim. Biophys. Acta 1131:333-336(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=95012454; PubMed=7927537;  
 RA Abdoet C.A., Baker E., Sutherland G.R., McCaughan G.W.;  
 RT "Genomic organization, exact localization, and tissue expression of the human CD26 (dipeptidyl peptidase IV) gene.";  
 RL Immunogenetics 40:331-338(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Peritoneal blood;  
 RX MEDLINE=92325476; PubMed=1352530;  
 RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,  
 RA Dahlberg H.N., Schlossman S.F., Morimoto C.;  
 RT "Cloning and functional expression of the T cell activation antigen CD26.";  
 RL J. Immunol. 149:481-486(1992).  
 RN [4]  
 RP ERRATUM.  
 RX MEDLINE=93171637; PubMed=8094732;  
 RA Tanaka T.;  
 RL J. Immunol. 150:2090-2090(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carinci P., Prance C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millady S.J.,  
 RA Boeck S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunatane P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 1-551 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=92165847; PubMed=1347043;  
 RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,  
 RA Trocort P., Barbat A.;  
 RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines HT-29 and Caco-2. Cloning of the complete human coding sequence and changes of dipeptidyl peptidase IV mRNA levels during cell differentiation.";  
 RL J. Biol. Chem. 267:4824-4833(1992).  
 RN [7]  
 RP SEQUENCE OF 545-766 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=91024044; PubMed=1977364;  
 RA Darmoul D., Lacasa M., Chantrel I., Swallow D., Trugnan G.;  
 RT "Isolation of a cDNA probe for the human intestinal dipeptidyl peptidase IV and assignment of the gene locus DPP4 to chromosome 2.";  
 RL Ann. Hum. Genet. 54:191-197(1990).  
 RN [8]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RX MEDLINE=96067599; PubMed=7487939;  
 RA Boehm S.K., Gum J.R., Jr., Erickson R.H., Hicks J.W., Kim Y.S.;  
 RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a TATA-less GC-rich sequence characteristic of a housekeeping gene promoter.";  
 RL Biochem. J. 311:835-843(1995).  
 RN [9]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Kidney;  
 RX MEDLINE=93210468; PubMed=8096237;  
 RA Morrison M.E., Vijayaradh S., Engelstein D., Albino A.P.,  
 RA Houghton A.N.;  
 RT "A marker for neoplastic progression of human melanocytes is a cell surface ectopeptidase.";  
 RL J. Exp. Med. 177:1135-1143(1993).  
 CC -I- FUNCTION: Removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. Plays a role in T cell activation.  
 CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-Xcc, from a polypeptide, preferentially when Xbb is Pro, provided Xcc is neither Pro nor hydroxyproline.  
 CC -I- SUBUNIT: Homodimer or heterodimer with seprase (FAP).  
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in a soluble form.  
 CC -I- PTM: The soluble form (SDPP) derives from the membrane form (MDPP) by proteolytic processing.  
 CC -I- SIMILARITY: Belongs to peptidase family S9B.  
 CC -I- DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd26.htm".  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)



RP REVISIONS.  
RA Marguet D.A.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B10.A; TISSUE=Liver;  
RX MEDLINE=95092780; PubMed=7999781;  
RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;  
RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene";  
RL Biochemistry 33:15204-15214(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schlier G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsien F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Piarne C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Rodriguez Y.S.N., Krzywinski M.I., Skalska U., Smolus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP SEQUENCE OF 1-20.  
RX MEDLINE=91302787; PubMed=1712807;  
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,  
RA Bernard A.-M., Gorvel J.-P., Pierres M.;  
RT "Evidence that thymocyte-activating molecule is mouse CD26  
(dipeptidyl peptidase IV)".  
RL J. Immunol. 147:447-454(1991).  
CC -I- FUNCTION: Removes N-terminal dipeptides sequentially from  
polypeptides having unsubstituted N-termini provided that the  
penultimate residue is proline.  
CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-  
Xcc, from a polypeptide, preferentially when Xbb is Pro, provided  
Xcc is neither Phe nor hydroxyproline.  
CC -I- SUBUNIT: Homodimer.  
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in  
a soluble form.  
CC -I- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)  
by proteolytic processing.  
CC -I- SIMILARITY: Belongs to peptidase family S9B.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL, X58384; CAA41274.1; -;  
DR EMBL, U12620; AAA82213.1; -;  
DR EMBL, U12599; AAA82213.1; JOINED.  
DR EMBL, U12600; AAA82213.1; JOINED.  
DR EMBL, U12601; AAA82213.1; JOINED.  
DR EMBL, U12602; AAA82213.1; JOINED.  
DR EMBL, U12603; AAA82213.1; JOINED.  
DR EMBL, U12604; AAA82213.1; JOINED.  
DR EMBL, U12605; AAA82213.1; JOINED.  
DR EMBL, U12606; AAA82213.1; JOINED.  
DR EMBL, U12607; AAA82213.1; JOINED.

DR EMBL, U12608; AAA82213.1; JOINED.  
DR EMBL, U12609; AAA82213.1; JOINED.  
DR EMBL, U12610; AAA82213.1; JOINED.  
DR EMBL, U12611; AAA82213.1; JOINED.  
DR EMBL, U12612; AAA82213.1; JOINED.  
DR EMBL, U12613; AAA82213.1; JOINED.  
DR EMBL, U12614; AAA82213.1; JOINED.  
DR EMBL, U12615; AAA82213.1; JOINED.  
DR EMBL, U12616; AAA82213.1; JOINED.  
DR EMBL, U12617; AAA82213.1; JOINED.  
DR EMBL, U12618; AAA82213.1; JOINED.  
DR EMBL, U12619; AAA82213.1; JOINED.  
DR EMBL, BC022183; AAH22183.1; -;  
DR MEROPS; S09.003; -;  
DR MGD; MGI:94919; Dpp4.  
DR InterPro; IPR002469; DppIV\_N\_term.  
DR InterPro; IPR001375; Peptidase\_S9.  
DR InterPro; IPR002471; Prol\_endopep\_ser.  
DR InterPro; IPR000379; Ser\_estra\_site.  
DR Pfam; PF00930; DppIV\_N\_term; 1.  
DR Pfam; PF00326; Peptidase\_S9; 1.  
DR PROSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
DR Hydrolase; AminoPeptidase; Dipeptidase; Serine protease;  
KW Transmembrane; Glycoprotein; Signal-anchor.  
FT CHAIN 1 760  
FT FT  
FT CHAIN 37 760  
FT FT  
FT DOMAIN 1 6  
FT TRANSMEM 7 28  
FT FT  
FT DOMAIN 29 760  
FT ACT\_SITE 295 466  
FT ACT\_SITE 624 624  
FT ACT\_SITE 702 702  
FT ACT\_SITE 734 734  
FT CARBOHYD 83 83  
FT CARBOHYD 90 90  
FT CARBOHYD 113 113  
FT CARBOHYD 213 213  
FT CARBOHYD 223 223  
FT CARBOHYD 315 315  
FT CARBOHYD 328 328  
FT CARBOHYD 514 514  
FT CARBOHYD 679 679  
FT SEQUENCE 760 AA; 87436 MW; ASF644B46A43DF8 CRC64;  
SQ  
Query Match 9.2%; Score 229; DB 1; Length 760;  
Best Local Similarity 22.6%; Pred. No. 1.5e-10;  
Matches 98; Conservative 47; Mismatches 120; Indels 168; Gaps 19;  
QY 140 KTGPHLYKITSILKSKYKRSSGGLPARSDPK-CPIKEIATITSGEWELGRHGSNIQV 198  
DB 372 KDGKHIHCHP-----PKDKDCTF-----ITGAWEMI-----SIA 403  
QY 199 DEVRLLVPEGT-KDSPLHLYVSYNPGVTRLTDRGYSHCIS-----QHCDFIT 252  
DB 404 LTSDLYYISNQYKMPGGRMLYKI-----QLTD--HTNVKCLSCDLNPERCOYVA 452  
QY 253 SKYSNQKPH---C-----VSLYKLSPEDDPTCKTKEFATILDSAG---PLPDTPP- 301  
DB 453 VSFKEAKYQUGCGPGPLPYTLTRSTDKELRYLE-----DNSALDRMLQDVOMPSK 506  
QY 302 --ISFSESTTGFTLYGMLYKPHDLOPKKKYPTLVFIYGP----- 339  
DB 507 KLDFIVLNEMTRF-WYOMILPH-PDKSKYPLLLDVAVGPGQKADASRLMWATYLAST 564  
QY 340 -----QVAIAG-- 345  
DB 565 ENIVASPDGRSGYGQDKIMAINRRLGTEVEDQIEAARQFYKMGFVDSKRVAIWGS 624  
QY 346 -----APVTLMIFDYDTGYTERYNG--HPDNOEGYLYGSYAM 380

Db 625 YGGYVTSMVLGSSGVFKCIGAVAPVSRMEYDVSYYTERYGLPIPEDNLHDYRNSTWNS 684  
 QY 381 QAEKFPSEPRRLLLHGFLENVFAHTSILLSFLVRAGKPYDLQIYPOERHSIRVPSG 440  
 Db 685 RAHF--KQVEYLHIGTADNVHFOQSAISALVDAGVDFOAMMYTDEDHGIASSTAN 742  
 QY 441 EHYELHLHYLQE 453  
 Db 743 QHYSHSHFLQ 755

RESULT 9  
 DPP6\_BOVIN  
 ID DPP6\_BOVIN STANDARD; PRT; 863 AA.  
 AC P42659;  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)  
 DE (Dipeptidyl aminopeptidase 6) (Dipeptidyl peptidase IV like protein)  
 GN (Dipeptidyl aminopeptidase-related protein) (DPPX).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NC NCB1\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=92108018; PubMed=1729689;  
 RA Wada K., Yokotani N., Hunter C., Doi K., Wentholt R.J., Shimazaki S.;  
 RT "Differential expression of two distinct forms of mRNA encoding  
 members of a dipeptidyl aminopeptidase family.";  
 Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).  
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN  
 ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF  
 ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE  
 FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=DPPX-L;  
 CC IsoId=P42659-1; Sequence=Displayed;  
 CC Name=DPPX-S;  
 CC IsoId=P42659-2; Sequence=VSP\_005364;  
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L  
 IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN  
 BRAIN, KIDNEY, OVARY AND TESTIS.  
 CC -!- SIMILARITY: Belongs to peptidase family S9B.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M76428; AAC1622.1; -  
 CC EMBL: M76429; AAC1623.1; -  
 CC PIR: A1793; A41793.  
 DR DR MEROPS: S09.973; -  
 DR InterPro: IPR002469; DPPV\_N\_term.  
 DR InterPro: IPR001375; Peptidase\_S9.  
 DR InterPro: IPR000379; Ser\_estr\_site.  
 DR Pfam: PF00930; DPPV\_N\_term.1.  
 DR Pfam: PF00326; Peptidase\_S9.1.  
 DR Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.  
 KM DOMAIN 1 93  
 FT TRANSMEM 94 114  
 FT DOMAIN 115 863  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT EXTRACELLULAR (POTENTIAL).

FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 811 811 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1 79 MASLYORFTGKINTSRSPFAPBASRLGGQEEDEGAPK  
 FT PLAGAOPAAPAPREPRGCGAGRPFOYARSCDDED ->  
 FT MTTAKERNASGKSVQOQEQ (In isoform DPPX-S).  
 FT /FTId=VSP\_005364.  
 SQ SEQUENCE 863 AA; 96556 MW; 23DBA792B841A39D CRC64;  
 Query Match 8.9%; Score 221; DB 1; Length 863;  
 Best local similarity 20.4%; Pred. No. 7,7e-10;  
 Matches 107; Conservative 65; Mismatches 178; Indels 174; Gaps 19;  
 QY 44 YIARAGWTPREGKYA--WSILDRSGTQIQLIISPELFIPEVDVNERQLIESVDSVT 101  
 Db 380 YITWYKMATSTVAVVM--LSRAQN-----VS-----LTLTCDATT 413  
 QY 102 PLII--YEETDIWINIHDFHVPQSHHEIEFIASECKTGFRLHYKITSILK--ESK 157  
 Db 414 GVTCKHDESEAWL-----HRONEEPVFS--KQG-RKFFVRALPQCGQCK 457  
 QY 158 YKRSSGGLPAPDFPCPIKEEIAITSGEWEVLGRHSNIOVEVRLVFEGTQDSPLEH 217  
 Db 456 FYHITVSSSQPRSSNDNIQ---SITSGMDVT---KILSYEKKSQIYFLSTEDLPRRR 510  
 QY 218 HLYVVSYNPGEVTR-----LTDGYSHSCCISGHCDFISKYSNQNKPCHVSLYKSS 271  
 Db 511 QLYSASTV--GSFNNQCUSCDLVNCTYFSASFSGAFLFKCGPGVP--TVSYNNTD 567  
 QY 272 PEDDPTCKTKE-FMATILDSAGLPDYPPELTFSESTGFLYGLMYLKHDLQGGKKVP 330  
 Db 566 KKGMDLEINENVQAISDRQPKVEYKIE-----TDVNLPIQLKPAFTDTATVP 621  
 QY 331 TVLFTYGGP-----  
 Db 622 LLLVVDGFGSGSAVEKFAVTWETVWSSHGAVVVKCDORGSGFGQTRLHEVRRRLGSL 681  
 QY 340 -----QVALAG-----APVT 349  
 Db 682 EEKDMEAVRWLKEPYIDKTRVAVFGDGYGLSTYLLPAKGDAQPVFSGSALSPIT 741  
 QY 350 LMIFYDTGTERYMGHPONEGYLLGSVAMQAEKFPSEPRRLLLHGFLENVFAHTS 409  
 Db 742 DPKLYASAFSEERYLDLGLDNRAEMAKVAVHRSAL--EGQQLVTHATADKTHFOHTA 799  
 QY 410 ILLSFLVRAGKPYDLQIYPOERHSIRVPSGHEYLHLHYLQE 453  
 Db 800 ELITQLIKKANYSLOIYFDSESHYSSAALQOHLHRSILGFEVE 843

RESULT 10  
 DPP6\_MOUSE  
 ID DPP6\_MOUSE STANDARD; PRT; 804 AA.  
 AC Q92218; Q9QW2; Q92219;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)  
 DE (Dipeptidyl aminopeptidase 6) (Dipeptidyl peptidase IV like protein)  
 GN (Dipeptidyl aminopeptidase-related protein) (DPPX).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Embryo;  
 RX MEDLINE=99030650; PubMed=9811881;  
 Hough R.B., Lengeling A., Bedian V., Lo C., Bucan M.;

RT "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-  
RT like protein 6 and causes dysregulation of kit expression.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).  
CC - FUNCTION: May be involved in the physiological processes of brain  
CC function. Has no dipeptidyl aminopeptidase activity. The lack of  
CC activity may be due to the substitution of an aspartate residue  
CC for the serine residue in the proposed catalytic triad (By  
CC similarity).  
CC - SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC - SIMILARITY: Belongs to peptidase family S9B.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
CC -----  
CC EMBL; AF092507; AAC97366.1; -;  
CC EMBL; AF092506; AAC97365.1; -;  
CC EMBL; AF092505; AAC98381.1; -;  
CC MEROPS; S09.973; -;  
CC MGD; MGI:94921; DPP6.  
CC InterPro; IPR002469; DPPV\_N term.  
CC InterPro; IPR001375; Peptidase\_S9.  
CC InterPro; IPR000379; Ser estersite.  
CC Pfam; PF00326; DPPV\_N term; 1.  
CC Pfam; PF00326; Peptidase\_S9; 1.  
CC TrasnMembrane; Glycoprotein; Signal-anchor.  
CC TRANSMEM 1 34 CYTOPLASMIC (POTENTIAL).  
CC FT DOMAIN 1 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
CC FT 35 55 (POTENTIAL).  
CC FT 56 804 EXTRACELLULAR (POTENTIAL).  
CC FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 440 445 MISSING (IN REF. 1; AAC98381).  
CC FT CONFLICT 638 638 O -> P (IN REF. 1; AAC97365).  
CC FT CONFLICT 804 AA; 91260 MM; 09CFCE7AADA8A7168 CRC64;  
CC SEQUENCE  
Query Match 8.8%; Score 218; DB 1; Length 804;  
Best Local Similarity 19.0%; Pred. No. 1.2e-09;  
Matches 110; Conservative 78; Mismatches 184; Indels 206; Gaps 22;  
QY 2 GTANPVTFTKMSIMIDAGRIIDVDKELIOFELFEGVEYIARAQMTPEKVA--WS 59  
DB 287 GSENPISLH---VGLNGPHTDL--EMPEPDDPRMREY-ITVWKAATSTKVAVTW- 337  
QY 60 ILIDBSQTLQIVLISPELFIPEDVMSRQRLIESVPSVPLLI--YEETDWINIH 117  
DB 338 --LNRAQN-----VS-----LILLCATTCVCKKHEDESEALH-- 370  
QY 118 DIFHHVPOGHEBEI-----EFTFASCKTGR-FLYKTSILKESKYRSSGGLPAPSD 170  
DB 371 -----RONEBPVPSKGRKFFVRAIPQGGGKGFVHT--VSSQPNSSNDNIQ---- 417  
QY 171 FKCPIKEELAIISGEMVELGRHSNIOVDEVRLLVFEQTSPLNHHIYVVSYNPGEV 230  
DB 418 -----SITSGMDVT---KILSYBEKRAKTYLSTEDLPRRRLHYSANTVD--DF 462  
QY 231 TR-----LTDGYSHSCISQHCDFISKYSNQNKPVC-----SLYKLSPEDDP 276  
DB 463 NROGLSCDLENCTVYASFSHMDPFLKCGSPGVPTVANTTDKRMFLLENANE--- 519  
QY 277 TCKTEFWATTILDSAGPLPDYTPPELFFSESTTGFTLLYMLYKPHDLQGGKYPVFLFY 336  
DB 520 -----EVOKAIRDQMPKIEYRKIEV-----EDYSLPMQIILKPATFTDTAYPLLLVVD 568

QY 337 GGP----- 339  
DB 569 GTPGSGSVTERFEVETMETVLVSSHGAVVVKCDRGSGFGQTKLLQEVRRLLGLEEKDQM 628  
QY 340 -----OVALNG-----APVTLWIFYD 355  
DB 629 EAVRTMLKGYIDKTRVAVFGKDYGLSTYILIPAKENGQGFTCGSALSPITDFLKYA 688  
QY 356 TGYTERYMGHPDQNEGGYVGVYVAMQAEKRPSPNNLLHGLFDEVNHPAHTSILSLFL 415  
DB 689 SASFSERYLGLHGDNDNAVENTKLARHVSAL--EDQFLIHTATDEKHPHTAEILITQL 746  
QY 416 VRAGKRPDIQIYQPERHSRVRPSSGHEYLHLHYOE 453  
DB 747 IKGKANYSLQIYPDSEHYHSAVKQHLRSRTITGFVE 784  
RESULT 11  
SEPR\_MOUSE  
ID SEPR\_MOUSE STANDARD; PRT; 761 AA.  
AC P97321;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Sequase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral  
DE membrane serine protease).  
GN PAP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10690;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RC STRAIN=BALB/c; TISSUE=Embryo;  
RX MEDLINE=9728459; PubMed=9139873;  
RA Niedermeier J., Scanlan M.J., Garin-Chesa P., Daiber C., Fiebig H.H.,  
RA Old L.J., Rettig W.J., Schnapp A.,  
RT "Mouse fibroblast activation protein: molecular cloning, alternative  
RT splicing and expression in the reactive stroma of epithelial  
RT cancer";  
RL Int. J. Cancer 71:383-389 (1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Breast;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buterow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M.J., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Bouffard G.G.,  
RA Bosa S.S., Loguellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Rask S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,  
RA Fahy U., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfeld Y.S.N., Krzywicki M.T., Skalska U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC - FUNCTION: May have a role in tissue remodeling during development  
CC and wound healing, and contribute to invasiveness in malignant  
CC cancers.  
CC - CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and  
CC type IV collagen, but not native type I or type IV collagen. Does  
CC not cleave laminin, fibronectin, fibrin or casein.  
CC - SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is





DR InterPro: IPR001375; Peptidase\_S9.  
 DR InterPro: IPR000379; Ser esters site.  
 DR Pfam: PF00930; DpIV N term; 1.  
 DR Pfam: PF00326; Peptidase\_S9; 1.  
 KM Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.  
 FT DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).  
 FT DOMAIN 117 865 EXTRACELLULAR (POTENTIAL).  
 FT CARBOHYD 173 173 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 319 319 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 404 404 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 471 471 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 535 535 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 566 566 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 813 813 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSPLIC 1 81 MASLYORFTCKINTSRFPAPPSHLLGGQCEDEGAGAKPAGPRAQAAAPRRGGGGGAGAGRPFGQGSDDGED  
 -> MTTAKERSASGKSVQOOEOE (in isoform DpIX-S).  
 FT FTId=VSP\_005365.  
 SQ SEQUENCE 865 AA; 97588 MW; 14B1AE0E024464B CRC64; .

Query Match 8.7%; Score 216; DB 1; Length 865;  
 Best Local Similarity 19.1%; Pred. No. 2e-09;

Matches 111; Conservative 79; Mismatches 179; Indels 212; Gaps 23;

QY 2 GTANPYTFKMSIMIDAEGRIIDVIDKELIOPELLFEVEVYIAAGWPEKKYA--WS 59  
 DB 348 GSNPFSISLH---VGLNCPHTDL--EMPPDDPRMREY-VITMKVATSTKVAVTW- 398  
 QY 60 ILIDRSQTRLQVLISPELLFIPVEDDVMERQRLIESVPSVPLII--YEETDIIINIH 117  
 DB 399 --LNRAQN-----VS-----ILTLCDATTGVCVKHDESEALH-- 431  
 QY 118 DIFHVPQSHSEET-----EIPFASCKTGFR-HLYKITSIKESKYRSSGGLPAPSD 170  
 DB 432 -----RONEEVFPKDKGKFFFIIRAIPQGGGKGFHIT--VSSSQPNSSNNIQ----- 478  
 QY 171 FKCPKEELAITGSEWEVLGRHGSNIQVDEVRLLVFEFGKDSPLHHLVYVSVNAPGEV 230  
 DB 479 -----SITSGMDVTV--KILAYDEKGNKITYFLSTEDLPRRRLYSAN----- 518  
 QY 231 TRLTDGRYSH---SCCISQCHDFPISKYSQKNPHCVSLYKLSPP-----EDDPTCKTK 281  
 DB 519 ---TEGNPFRQGLSCDLVENCYTFASFSHSM---FLLKCGCPGVPMVYVNTITDKK 572  
 QY 282 EF-----WATILDSAGPLPYTPPEIFSESTGFTLYGMLYKPHDLQPGKXYPTVL 333  
 DB 573 MEDLETNEHVKKAINDRQPKVEYRDIET-----DDVNLPMQILKPAFTDTTHYPLL 626  
 QY 334 FIYGGP----- 339  
 DB 627 VVDGTFGSSVAEKEFVSEVETWVSHGAVVVKCDGRSGGPGCKLLHEVRRLGLEEK 686  
 QY 340 -----OVAIAG-----APVTLWI 352  
 DB 687 DQMEAVRTMLKEQYIDRTVAVAGKDYGYLSTYILPAKENGSGQFTFGSALSPITDKK 746  
 QY 353 FPDGTGTERYMGHPDNEGGYLYGSVAMQAEKPPSPNNLLHGLDENVHPAFTSILL 412  
 DB 747 LVASAFSEERYLGHGIDNRAVEYTKVAHVSAL--EEQOFLIHPADEKIHFOHTAELI 804  
 QY 413 SFVLVAGKPYDQIYPOERHSIRVPESGEYELHLLHYOE 453  
 DB 805 TOLIRKANYSLQIYDESHYFTSSSLKQHLRSIINFVE 845

RESULT 13  
 SEPR\_HUMAN STANDARD: PRT; 760 AA.  
 AC Q12884; 000199; Q99998; Q9UD4;  
 DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Sepsinase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral  
 DE membrane serine protease) (170-kDa melanoma membrane-bound  
 DE gelatinase).  
 GN FAP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RP TISSUE=Fibroblast;  
 RX MEDLINE=94261645; PubMed=7911242;  
 RA Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncali M.P.,  
 RT Healey J.H., Old L.J., Rettig W.J.;  
 RT "Molecular cloning of fibroblast activation protein alpha, a member of  
 RT the serine protease family selectively expressed in stromal  
 RT fibroblasts of epithelial cancers";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661 (1994).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RP TISSUE=Melanoma;  
 RX MEDLINE=97388251; PubMed=9247085;  
 RA Goldstein L.A., Ghersi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y.,  
 RA Plesgate D., Chen W.-T.;  
 RT "Molecular cloning of seprase: a serine integral membrane protease  
 RT from human melanoma";  
 RL Biochim. Biophys. Acta 1361:11-19 (1997).  
 RN (3)  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND  
 RP 511-518.  
 RP TISSUE=Melanoma;  
 RX MEDLINE=97218181; PubMed=9065413;  
 RA Pineiro-Sanchez M.L., Goldstein L.A., Dodd J., Howard L., Yeh Y.,  
 RA Chen W.-T.;  
 RT "Identification of the 170-kDa melanoma membrane-bound gelatinase  
 RT (seprase) as a serine integral membrane protease";  
 RL J. Biol. Chem. 272:7595-7601 (1997).  
 RN (4)  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RP TISSUE=Melanoma;  
 RX MEDLINE=20112818; PubMed=10644713;  
 RA Goldstein L.A., Chen W.-T.;  
 RT "Identification of an alternatively spliced seprase mRNA that encodes  
 RT a novel intracellular isoform";  
 RL J. Biol. Chem. 275:2554-2559 (2000).  
 RN (5)  
 RP SEQUENCE OF 192-208; 220-240 AND 510-521.  
 RX MEDLINE=94327249; PubMed=7519584;  
 RA Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,  
 RA Garin-Chesa P., Healey J.H., Old L.J.;  
 RT "Fibroblast activation protein: purification, epitope mapping and  
 RT induction by growth factors";  
 RL Int. J. Cancer 58:385-392 (1994).  
 CC -!- FUNCTION: May have a role in tissue remodeling during development  
 CC and wound healing, and may contribute to invasiveness in malignant  
 CC cancers.  
 CC -!- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and  
 CC type IV collagen, but not native type I or type IV collagen. Does  
 CC not cleave laminin, fibronectin, fibrin or casein.  
 CC -!- SUBUNIT: Homodimer, or heterodimer with DP4. The monomer is  
 CC inactive.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell  
 CC surface lamellipodia, invadopodia and on shed vesicles.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=L;  
 CC IsoId=Q12884-1; Sequence=Displayed;  
 CC Note=Major isoform;  
 CC Name=2; Synonyms=S, Truncated;  
 CC IsoId=Q12884-2; Sequence=VSP\_005367;  
 CC -!- TISSUE SPECIFICITY: Fibroblast-specific.

CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling  
CC during development, tissue repair, and carcinogenesis.  
CC -1- PTM: N-glycosylated.  
CC -1- SIMILARITY: Belongs to peptidase family S9B.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U09278; AAB49652.1; -  
CC EMBL; U76833; AAC51668.1; -  
CC EMBL; AF007822; AAF21600.1; -  
CC MEROPS; S09\_007; -  
CC Genew; HGNC:3590; PAP.  
CC MIM; 600403; -  
CC InterPro; IPR002469; DPPIV N term.  
CC InterPro; IPR001375; Peptidase S9.  
CC InterPro; IPR002471; Prol endopep ser.  
CC InterPro; IPR000379; Ser\_estrs\_site.  
CC Pfam; PF00930; DPPIV\_N\_term; 1.  
CC PROSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
CC PROSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
CC Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;  
CC Glycoprotein; Alternative splicing.  
CC  
CC DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
CC (POTENTIAL).  
CC  
CC DOMAIN 26 760 EXTRACELLULAR (POTENTIAL).  
CC ACT SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).  
CC ACT\_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).  
CC ACT\_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).  
CC CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1 521 Missing (in isoform 2).  
CC VARSPLIC /Prtid=VSP\_005367.  
CC  
CC CONFLICT 207 207 P -> A (IN REF. 2).  
CC CONFLICT 229 229 K -> T (IN REF. 2).  
CC CONFLICT 354 354 R -> T (IN REF. 2).  
CC  
CC SEQUENCE 760 AA; 87820 MW; A0D34B4801B07EA CRC64;  
SQ  
Query Match 8.64; Score 214; DB 1; Length 760;  
Best Local Similarity 20.24; Pred. No. 2,4e-09;  
Matches 87; Conservative 53; Mismatches 127; Indels 164; Gaps 13;  
QY 140 KTGFRHLKYKTSILKESKYSRSGGLPAPDFKCPKEEIAITSGEVEVGRHSNQVD 139  
DB 371 KQGYKIHVI-----KDYVNAIQITSGKKEAI-----NI--- 400  
QY 200 EVARLVFECTKDSPLEHLLVYVYVNGEVTRITDRGYSHS--CCISQH-----CDFIS 253  
DB 401 -----FRVTQOSLFYSSNEFEYPPGRNIRIYISIGSYPPSKVCVTCILRKRCQYTA 453  
QY 254 KYSNQKRPCHVSLY---KLSPEDDPTCKTEFWAITLDSAGPLD-----YTPPEIFS 304  
DB 454 SFSDYAKYVALVYGGIPISLHDTGTDG---IKILENKELENALNQLPKEEIK 509  
QY 305 FEFTTFTLYGMLYKPHDLQPKKKYPTLVFIYGGP----- 339  
DB 510 KLEVDETILMYKNILPPQPRSKKYPILLIQVYGGPCQSQSVRSFVAVVMISYLSKSGMVI 569  
QY 340 ----- 339  
DB 570 ALVDGRGTAFOGDKLLYAVYRKLGVEVEDOIYAVRKEIFEMGFIDEKRIAIMGMSYGYV 629  
QY 340 -----QVALAGAVTLMIFFDTGYTERVMGHP--DQNEQGYLYGVAMQAEKF 385

DB 630 SSLALASGTGLPFCGIAVAVPVSSWEYVASVYTERFNGLPKCDNLHYNKSTWARRAEYF 689  
QY 386 PSEPNRLLLHGFIDENVAFAHTSILSLFVYRAGKAYDIOIYQERHSIRVPESG---EH 442  
DB 690 RAVD--YLLHGHADNVAIFQNSAQAKALVNAQVFOAMWYSDQNHGL-----SGLSLNH 743  
QY 443 YELHLHYLQOE 453  
DB 744 LYTHMTHFLKQ 754  
RESULT 14  
ID DPP6\_RAT STANDARD; PRT; 859 AA.  
AC P46101;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)  
DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)  
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).  
GN DPP6  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).  
RC TISSUE=Brain;  
RX MEDLINE=92108018; PubMed=1729689;  
RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimaaki S.;  
RT "Differential expression of two distinct forms of mRNA encoding  
RT members of a dipeptidyl aminopeptidase family."  
RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).  
CC -1- FUNCTION: May be involved in the physiological processes of brain  
CC function. Has no dipeptidyl aminopeptidase activity. The lack of  
CC activity may be due to the substitution of an aspartate residue  
CC for the serine residue in the proposed catalytic triad.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=DPPX-L;  
CC IsoId=P46101-1; Sequence=Displayed;  
CC Name=DPPX-S;  
CC IsoId=P46101-2; Sequence=VSP\_005366;  
CC TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME  
CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS. IN  
CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.  
CC -1- SIMILARITY: Belongs to peptidase family S9B.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M76426; AAC42061.1; -  
CC EMBL; M76427; AAC42062.1; -  
CC MEROPS; S09\_973; -  
CC InterPro; IPR002469; DPPIV N term.  
CC InterPro; IPR001375; Peptidase S9.  
CC InterPro; IPR000379; Ser\_estrs\_site.  
CC Pfam; PF00930; DPPIV\_N\_term; 1.  
CC Pfam; PF00326; Peptidase\_S9; 1.  
CC Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.  
CC  
CC DOMAIN 1 89 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 90 110 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
CC (POTENTIAL).  
CC  
CC DOMAIN 111 859 EXTRACELLULAR (POTENTIAL).  
CC CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT 0

```

FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 75 MASLYRFTKINTSRSPAPRASHLLGGGSPEDAGSKP
      LGPOAAVAPRERGAGGRPRFOYARSDCCDED -> MTT
      AKERSASGKSVOOODO (in isoform DPPX-S).
      /FTID=VSP_005366.
SQ SEQUENCE 859 AA; 97301 MW; CE26856D26ED1268 CRC64;

Query Match 8.6%; Score 214; DB 1; Length 859;
Best Local Similarity 18.9%; Pred. No. 2.8e-09;
Matches 109; Conservative 81; Mismatches 182; Indels 206; Gaps 22;

QY 2 GTANPKYTFKMSIIMDAEGRIDVIDKELIQPEILFEVEYIARAGWTEGKYA-WS 59
   342 GSENPISILH---VIGLNGPTHDL---EMMPDDPRMRX-YITVMKATSTKVAIVM- 392
QY 60 ILDRSQRLOIYLISPELFIPEDDWDERORLIESVDSVTPILI--YEETDIWINIH 117
   393 --LNRQN-----VS-----LITLCDATGYCTKKEHDESSAMLA-- 425
QY 118 DIFHVPQSHBEI-----EFIFASECKTGFR-HLYKITSILKESKYKSGGLPAPSD 170
   426 -----RQNEPFPKSDGRKFVFRAPQGRGKFTIIT--VSSQPNSSNDNIQ---- 472
QY 171 FKCPIKEIATSGSEWEVLGRHGSNIQVDEVRLVYFEGTQSPLEHLVYVSVNPGEV 230
   473 -----SITSGDMVLT-----ELITYDEKRNKLTFLSTEDLPRRRLVYANVTDDNR 519
QY 231 TRLTDRGYSHCCTSCQCHDFISKYSNOKMHCYSLYLSP-----EDPCTKTEF- 283
   520 QCL-----SCDLVENCITYVASFSNMD---FLLKCEBPGVPTVTHNTTKRRKRFD 569
QY 284 -----WATLDSAGPLDVTYTPPEIFSESTGFTLYGMYYKPHDLQPKKYPTVLFYI 336
   570 LEANEQYKAIYDQMKEIYRKIEV-----EDYSLPMQLKATFTDTAHYELLYVD 623
QY 337 GGP----- 339
   624 GTPGSGVSERFEVETWETVLVSHGAVVVKCDGRSGFGQTKLHEVRRRLGLEBKDM 683
QY 340 -----QVALAG-----APVLMIFYD 355
   684 EAVRTMLKEOYIDKTRVAVFGKDYGGYLSTYILPAKGENOGQFTCSALSPIDFKLYA 743
QY 356 TGYTRYMGHPDQNEGGYLLGSVAMQAKFPSEPRRLLLHGFDEVNHFHTSILLSFL 415
   744 SASERYLGLHLDNRAYEMTKLAHRSAL--EDQQLIHAATADKIHFOHIAELITQL 801
QY 416 VRACKPYDLOIYQERHSIRVSEGEHYELHLYLOE 453
   802 IKGKANYSLQIYPDESHYFHSVALKQHLXRSIIIGFVE 839
Db

RESULT 15
YDZF SCHPO STANDARD; PRT; 853 AA.
AC 09P7E9; 013720;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DN Putative dipeptidyl aminopeptidase Cl4C4.15c (EC 3.4.14.-).
GN SPAC14C4.15C OR SPAPJ760.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Schizosaccharomycetes; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_taxid=4896;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajendram M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Heitroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Shaper K.,
RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Glymptre B.,
RA Welljans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert A.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Mambitt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nure P.,
RT "the genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
-1 SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
  vacuoles.
-1 SIMILARITY: Belongs to peptidase family S9B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AL162631; CAB83084.1; -
DR EMBL; Z98596; CAB1208.1; -
DR GenDB; Spombe; SPAC14C4.15C; -
DR InterPro; IPR002459; DPPIV N term.
DR InterPro; IPR001375; Peptidase-S9.
DR Pfam; PF00930; DPPIV N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KW Serineprotease; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 66 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 67 89 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT LUMENAL (POTENTIAL).
FT ACT_SITE 719 853 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 795 795 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 828 828 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 853 AA; 98341 MW; 38450BA50F8304B6 CRC64;

Query Match 9.9%; Score 145.5; DB 1; Length 853;
Best Local Similarity 19.0%; Pred. No. 0.00091;
Matches 116; Conservative 65; Mismatches 179; Indels 249; Gaps 25;

QY 26 VIDEKLIQPEILFEVEYIARAGWTEGKYAMSILDRSQRLOIYLISPELFIPEDD 85
   277 YVEEIIQSSKAVM-----WSPDGNCLSYLSIDSKVPVH-VLPEQDLSKVEDQ 325

```

```
OY 86 VMERQRLIESVDPSTPLIYEETDOWINIH-----DIPHPQOS--HEEIEFIAPA 136
DB 326 NRVNPFHYSTPKDIPFV-----KLFVNCFTDGESEIEVDSFPLSTQHRXYITDVAMA 379
OY 137 SECKTGF-----RHLY-KITSILKESKYKSSGGLPAPSDPKPIKEEIAITSG-----E 185
DB 380 GNEKLMFVEVLNGNERYVTSLPDLSRKTTIENTEVS-----EHPIALTSSLHKYILS 432
OY 186 WEVLGRHGSNIQVDEVRRLVPEGTK-----DSPLEHMLYVVSYYNPGEVTRLTD-- 235
DB 433 FESLIG---NLKERYVRQ--YFLSNKKRIAIYELDNF-----VPILYLPVNISFLSDLY 480
OY 236 -----RGYSHSCCISQ-----HCFPIISKYSNOKNPHCVS---- 265
DB 481 LINTLYFTAISGSPFSRVRLCTKSILISEINIOIGSGLFGIKVSNDOYILVNYLGP 540
OY 266 -----LY-----KLSPEDDPT-----CK-----TKEFWA 285
DB 541 EIPKOFIYSIHEDKVSTSNDSKNNLPDSSTSLGKVKLELCSLETNEELIITKEKFA 600
OY 286 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFYGGPOVA--- 342
DB 601 -----FPGVFFKVIKVNKITAYIQEIRPENPRKRYPTVPHLYGAPQALVT 648
OY 343 -----IAG----- 345
DB 649 GKEYMDINELMASVYNFLVIVKVIDIDSVSGOHLFSDSHELLIKSMIELLSRYVDTPEYI 708
OY 346 -----APVTWIEYDTGYTERYMGHPDQNEOG 372
DB 709 DRHRVGIWGSFGGYLTLLKILENADFIKTGAVALPAPTDWRYDYDAYISENLLGAYSKQTTA 768
OY 373 YVLGSVAMQAEKFPSEPNRLLLHGFLDENVHFAHTSILLSFLVRAG---KPYDLOIYP 428
DB 769 IYDKTAHYHSENF-RKLGGLVLVHGTSDDNVHIENMTQLTAMVEKGVYNYVPF---IVP 824
OY 429 QERHSIRVP 437
DB 825 NANHEFSDP 833
```

Search completed: October 15, 2003, 17:11:32  
Job time : 11.3773 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 | Search time 29.7397 Seconds

(without alignments)  
4034.822 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKVTFKMSIMDAE.....HLHYIGENGSRIALAKVI 465

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SPTREMBL\_23:\*

- 1: sp\_archea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2482	100.0	465	Q9HBM3	Q9hbm3 homo sapien
2	2447.5	98.6	831	Q9NEM5	Q9nem5 homo sapien
3	2441.5	98.4	632	Q96UX1	Q96ux1 homo sapien
4	2422	97.6	882	Q9HBM5	Q9hbm5 homo sapien
5	2422	97.6	882	Q8IMG7	Q8img7 homo sapien
6	2326	93.7	892	Q9DAG6	Q9dag6 mus musculu
7	1494.5	60.2	863	Q8WYD8	Q8wyd8 homo sapien
8	1488.5	60.0	862	Q8BWT9	Q8bwt9 mus musculu
9	1478.5	59.6	862	Q8BVG4	Q8bvg4 mus musculu
10	1377.5	55.5	310	Q9HBM4	Q9hbm4 mus sapien
11	1266.5	51.0	628	Q9N3F5	Q9n3f5 homo sapien
12	1187.5	47.8	508	Q75273	Q75273 homo sapien
13	1154.5	46.5	439	Q9BVR3	Q9bvr3 homo sapien
14	860.5	34.7	360	Q9HBM2	Q9hbm2 homo sapien
15	845	34.0	1042	Q9VC20	Q9vc20 drosophila
16	845	34.0	1102	Q9VC19	Q9vc19 drosophila

17	841.5	33.9	1053	Q8IH07	Q8ih07 drosophila
18	787	31.7	469	Q9NXP4	Q9nxf4 homo sapien
19	719	29.0	312	Q96NT8	Q96nt8 homo sapien
20	474.5	19.1	167	Q9N2J7	Q9n2j7 homo sapien
21	462	18.6	927	Q965K3	Q965k3 caenorhabdi
22	462	18.6	931	Q44987	Q44987 caenorhabdi
23	449	18.1	552	Q8GJY7	Q8gjy7 arabidopsis
24	449	18.1	746	Q9FNF6	Q9fnf6 arabidopsis
25	429.5	17.3	728	Q9A6E0	Q9a6e0 caulobacter
26	415.5	16.7	763	Q8EAB7	Q8eab7 shewanella
27	355.5	14.3	751	Q8P3V8	Q8p3v8 xanthomonas
28	355	14.3	741	P95782	P95782 xanthomonas
29	346	13.9	757	Q8PFD7	Q8pfd7 xanthomonas
30	312	12.6	901	Q96VT7	Q96vt7 aspergillus
31	309.5	12.5	771	Q42812	Q42812 aspergillus
32	308.5	12.4	765	Q14425	Q14425 aspergillus
33	305	12.3	723	Q31048	Q31048 porphyromon
34	305	12.3	723	Q66223	Q66223 porphyromon
35	300.5	12.1	711	Q47900	Q47900 flavobacter
36	263.5	10.6	730	Q93JY4	Q93jy4 prevotella
37	263	10.6	748	P70092	P70092 xenopus lae
38	247	10.0	711	Q9P236	Q9p236 homo sapien
39	239	9.6	796	Q8N608	Q8n608 homo sapien
40	234	9.4	827	Q8PPU4	Q8ppu4 xanthomonas
41	227	9.1	761	Q8R492	Q8r492 rattus norv
42	226	9.1	799	Q18119	Q18119 caenorhabdi
43	225.5	9.1	803	Q8P500	Q8p500 xanthomonas
44	224.5	9.0	755	Q91651	Q91651 xenopus lae
45	222.5	9.0	745	Q9VMM2	Q9vmm2 drosophila

## ALIGNMENTS

## RESULT 1

ID	Q9HBM3	PRELIMINARY;	PRT;	465 AA.
AC	Q9HBM3			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Dipeptidyl peptidase 8 (Fragment).			
DN	DP8.			
OS	Homo sapiens (human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=20467194; PubMed=11012666;			
RA	Abdott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,			
RA	Gorell M.D.;			
RT	"Cloning, expression and chromosomal localization of a novel human			
RT	dipeptidyl peptidase (DPP) IV homolog, DPP8."			
RL	Eur. J. Biochem. 267:6140-6150(2000).			
DR	EMBL; AF221636; AAC29768.1; -			
DR	MEROPS; S09_016; -			
FT	NON TER			
FT	1			
SQ	SEQUENCE 465 AA; 53197 MW; 229399EC0A4FE29CE CRC64;			

Query Match 100.0%; Score 2482; DB 4; Length 465;  
Best Local Similarity 100.0%; Pred. No. 6.1e-207;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TGTANPKVTFKMSIMDAEGRRIIDVIDKELIDPFEILFGVEYIRAGATPGKYAWSI 60
Db	1	TGTANPKVTFKMSIMDAEGRRIIDVIDKELIDPFEILFGVEYIRAGATPGKYAWSI 60
Qy	61	LDRSQTRLQIVLISEPLFIPVEDDVMERQRLIESVPSVTPPLIYEETDWINIHDF 120
Db	61	LDRSQTRLQIVLISEPLFIPVEDDVMERQRLIESVPSVTPPLIYEETDWINIHDF 120

```
QY 121 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 180
DB 121 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 180
QY 181 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 240
DB 181 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 240
QY 241 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 300
DB 241 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 300
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVAIAGAPVTLMIFDGTGYTE 360
DB 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVAIAGAPVTLMIFDGTGYTE 360
QY 361 RYMGHPONEGYYLGSVAMQAEKPPSEPNRLLLHGLDENVFAHTSILSLFVYRAGK 420
DB 361 RYMGHPONEGYYLGSVAMQAEKPPSEPNRLLLHGLDENVFAHTSILSLFVYRAGK 420
QY 421 PYDLQIYPOERHSIRVPESGHEYLHLHYLOENLGSRIALAKVI 465
DB 421 PYDLQIYPOERHSIRVPESGHEYLHLHYLOENLGSRIALAKVI 465
```

## RESULT 2

```
Q8NEM5 PRELIMINARY; PRT; 831 AA.
ID Q8NEM5
AC Q8NEM5;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Similar to dipeptidylpeptidase 8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strauberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC030688; AAH30688.1; -
SQ SEQUENCE 831 AA; 95527 MW; 0B2A13A2FE70CB2 CRC64;
```

Query Match 98.6%; Score 2447.5; DB 4; Length 831;  
Best Local Similarity 90.5%; Pred. No. 1.4e-203;  
Matches 465; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```
QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFLLFEGVEYIARAGWTPREGKXWMSI 60
DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFLLFEGVEYIARAGWTPREGKXWMSI 377
QY 61 LLDRSQTRLQIVLISPELFIPEDDVMERQRLIESVPDSVTPPLIYEETDIDIMINHDIF 120
DB 378 LLDRSQTRLQIVLISPELFIPEDDVMERQRLIESVPDSVTPPLIYEETDIDIMINHDIF 437
QY 121 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 180
DB 121 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 180
QY 438 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 497
DB 438 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 497
QY 181 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 240
DB 181 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 240
QY 498 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 557
DB 498 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 557
QY 241 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 300
DB 241 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 300
QY 558 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 617
DB 558 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 617
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPO----- 340
DB 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVOLVNNRFGKVKYFRNLTLA 677
```

```
QY 341 -----VAIAGAPVTLMIFDGTGYTERYMGHPDNEQ 371
DB 678 SLGVVVVVDNRSGCHRGKLFEGAFKYKQVAIAGAPVTLMIFDGTGYTERYMGHPDNEQ 737
QY 372 GYVLGSVAMQAEKPPSEPNRLLLHGLDENVFAHTSILSLFVYRAGKPYDLQIYPOER 431
DB 738 GYVLGSVAMQAEKPPSEPNRLLLHGLDENVFAHTSILSLFVYRAGKPYDLQIYPOER 797
QY 432 HSIRVPESGHEYLHLHYLOENLGSRIALAKVI 465
DB 798 HSIRVPESGHEYLHLHYLOENLGSRIALAKVI 831
```

## RESULT 3

```
Q96JX1 PRELIMINARY; PRT; 632 AA.
ID Q96JX1
AC Q96JX1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein FLJ14920.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Iwagaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Makamatsu A., Nakamura Y., Nagahari K., Maeno Y.,
RA Ninomiya K., Iwanagaki T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027826; BAB55395.1; -
DR MEROPS; S09.018; -
KM Hypothetical protein.
SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;
```

Query Match 98.4%; Score 2441.5; DB 4; Length 632;  
Best Local Similarity 90.3%; Pred. No. 3.1e-203;  
Matches 484; Conservative 0; Mismatches 1; Indels 49; Gaps 1;

```
QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFLLFEGVEYIARAGWTPREGKXWMSI 60
DB 119 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFLLFEGVEYIARAGWTPREGKXWMSI 178
QY 61 LLDRSQTRLQIVLISPELFIPEDDVMERQRLIESVPDSVTPPLIYEETDIDIMINHDIF 120
DB 179 LLDRSQTRLQIVLISPELFIPEDDVMERQRLIESVPDSVTPPLIYEETDIDIMINHDIF 238
QY 121 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 180
DB 121 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 180
QY 239 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 298
DB 239 HVFPOSHBEIEIFASECKTGFRHLKYITSIILKESYKXSSGGLPAPSPDFCKPIKEIA 298
QY 181 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 240
DB 181 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 240
QY 299 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 358
DB 299 ITSGEMVLGRHGSNIQVDEVRRLVFEGETKDSPLFHLHYVSVVNGEYTRLLDRGYSH 358
QY 241 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 300
DB 241 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 300
QY 359 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 418
DB 359 SCCISQHCDFEISKYSNQKNPHCVSLYKLSPPEDPTCKTEFWATILDSAGPLPDYTPP 418
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPO----- 340
DB 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVOLVNNRFGKVKYFRNLTLA 478
QY 341 -----VAIAGAPVTLMIFDGTGYTERYMGHPDNEQ 371
DB 479 SLGVVVVVDNRSGCHRGKLFEGAFKYKQVAIAGAPVTLMIFDGTGYTERYMGHPDNEQ 538
```

QY 372 GYLLGSVAMQAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQER 431  
DB 539 GYLLGSVAMQAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQER 538  
QY 432 HSIRVPESGEYELHLHLHYQENLGSRIALAKVI 465  
DB 599 HSIRVPESGEYELHLHLHYQENLGSRIALAKVI 632

RESULT 4  
Q9HBM5 PRELIMINARY; PRT; 882 AA.  
ID Q9HBM5  
AC Q9HBM5  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Dipeptidyl peptidase 8.  
GN DPP8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20467194; Pubmed=11012666;  
RA Abbott C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,  
RA Gorrell M.D.,  
RT "Cloning, expression and chromosomal localization of a novel human  
RT dipeptidyl peptidase (DPP) IV homology, DPP8".  
RL Eur. J. Biochem. 267:6140-6150 (2000).  
DR EMBL: AF221634; AAG29766.1; -  
DR MEROPS: S09.018; -  
DR InterPro: IPR002469; DPPV\_N\_term.  
DR InterPro: IPR001375; Peptidase\_S9.  
DR InterPro: IPR000379; Ser\_estrs\_site.  
DR Pfam: PF00930; DPPV\_N\_term; 1.  
DR Pfam: PF00326; Peptidase\_S9; 1.  
DR SEQUENCE 882 AA; 101421 MW; AD801C302DB46528 CRC64;

Query Match 97.6%; Score 2422; DB 4; Length 882;  
Best Local Similarity 82.3%; Pred. No. 2.5e-201;  
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKSEIMIDAGRIIDYIDKELIQPEILFEGVEYIARAGTPEGKAWSI 60  
DB 318 TGTANPKVTFKSEIMIDAGRIIDYIDKELIQPEILFEGVEYIARAGTPEGKAWSI 377  
QY 61 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPSVTPPLIYEETTDIWINIDIF 120  
DB 378 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPSVTPPLIYEETTDIWINIDIF 437  
QY 121 HVPQSHHEIEIFIFASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEIA 180  
DB 438 HVPQSHHEIEIFIFASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEIA 497  
QY 181 ITSGEVEVLRGHSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGVETRLTDGYSH 240  
DB 498 ITSGEVEVLRGHSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGVETRLTDGYSH 557  
QY 241 SCCISQHCDFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGLPDYTPP 300  
DB 558 SCCISQHCDFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGLPDYTPP 617  
QY 301 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQ----- 340  
DB 618 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFKGVKXFRILNTLA 677  
QY 341 -----VAIAGAPVTLMIFFYDTGTYTERYMGHPDQNEGYLLGSVAM 340  
DB 678 SLGYYVVVVDNRGSGHGLKFEAGAFKXKMGQIEIDQVEGLQYLASRYDIFDLDRVGIHG 737  
QY 341 -----VAIAGAPVTLMIFFYDTGTYTERYMGHPDQNEGYLLGSVAM 380

DB 738 WSYGYLISLALMQRSDIFRVALIAGAPVTLMIFFYDTGTYTERYMGHPDQNEGYLLGSVAM 797  
QY 381 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 440  
DB 798 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 857  
QY 441 EHYELHLHLHYQENLGSRIALAKVI 465  
DB 858 EHYELHLHLHYQENLGSRIALAKVI 882

RESULT 5  
Q8IMG7 PRELIMINARY; PRT; 882 AA.  
ID Q8IMG7  
AC Q8IMG7  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Similar to dipeptidyl peptidase 8.  
GN  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC040203; AAH40203.1; -  
DR SEQUENCE 882 AA; 101391 MW; 88C76AF5BCE707F9 CRC64;

Query Match 97.6%; Score 2422; DB 4; Length 882;  
Best Local Similarity 82.3%; Pred. No. 2.5e-201;  
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKSEIMIDAGRIIDYIDKELIQPEILFEGVEYIARAGTPEGKAWSI 60  
DB 318 TGTANPKVTFKSEIMIDAGRIIDYIDKELIQPEILFEGVEYIARAGTPEGKAWSI 377  
QY 61 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPSVTPPLIYEETTDIWINIDIF 120  
DB 378 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPSVTPPLIYEETTDIWINIDIF 437  
QY 121 HVPQSHHEIEIFIFASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEIA 180  
DB 438 HVPQSHHEIEIFIFASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEIA 497  
QY 181 ITSGEVEVLRGHSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGVETRLTDGYSH 240  
DB 498 ITSGEVEVLRGHSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGVETRLTDGYSH 557  
QY 241 SCCISQHCDFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGLPDYTPP 300  
DB 558 SCCISQHCDFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGLPDYTPP 617  
QY 301 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQ----- 340  
DB 618 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFKGVKXFRILNTLA 677  
QY 341 -----VAIAGAPVTLMIFFYDTGTYTERYMGHPDQNEGYLLGSVAM 340  
DB 678 SLGYYVVVVDNRGSGHGLKFEAGAFKXKMGQIEIDQVEGLQYLASRYDIFDLDRVGIHG 737  
QY 341 -----VAIAGAPVTLMIFFYDTGTYTERYMGHPDQNEGYLLGSVAM 380  
DB 728 WSYGYLISLALMQRSDIFRVALIAGAPVTLMIFFYDTGTYTERYMGHPDQNEGYLLGSVAM 797  
QY 381 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 440  
DB 798 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 857  
QY 441 EHYELHLHLHYQENLGSRIALAKVI 465



Db 858 EHVELHHTYGENICSRIALAKVI 882

RESULT 6

09D4G6 PRELIMINARY; PRT; 892 AA.

AC 09D4G6

AD 09D4G6:

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Adult male testis cDNA, RIKEN full-length enriched library,

DE clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=CS7BL/6J; TISSUE=Testis;

RC Adachi J., Aizawa K., Akhita S., Akimura T., Arai A., Aono H.,

RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RN SEQUENCE FROM N.A.

RP STRAIN=CS7BL/6J; TISSUE=Testis;

RP MEDLINE=22354683; PubMed=12466851;

RX The FANTOM Consortium

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

RN [3]

RN SEQUENCE FROM N.A.

RP STRAIN=CS7BL/6J; TISSUE=Testis;

RP MEDLINE=21085660; PubMed=11217851;

RA RIKEN FANTOM Consortium;

RT "Functional annotation of a full-length mouse cDNA collection.";

RT Nature 409:685-690(2001).

RN [4]

RN SEQUENCE FROM N.A.

RC STRAIN=CS7BL/6J; TISSUE=Testis;

RC MEDLINE=99279253; PubMed=10349636;

RA Carninci P., Hayashizaki Y.;

RT "High-efficiency full-length cDNA cloning.";

RL Meth. Enzymol. 303:19-44(1999).

RN [5]

RN SEQUENCE FROM N.A.

RC STRAIN=CS7BL/6J; TISSUE=Testis;

RC MEDLINE=20499374; PubMed=11042159;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

RT "Normalisation and subtraction of cap-trapper-selected cDNAs to

RT prepare full-length cDNA libraries for rapid discovery of new genes.";

RL Genome Res. 10:1617-1630(2000).

RL [6]

RL SEQUENCE FROM N.A.

RC STRAIN=CS7BL/6J; TISSUE=Testis;

RC MEDLINE=20530913; PubMed=11076861;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Saeki N., Carninci P.,

RA Kono H., Aiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,

RA Sumi N., Ishii Y., Nakamura S., Harada M., Nishino T., Harada A.,

RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujiwake S., Inoue K., Togawa Y., Izawa S., Ohata E., Watanaki M.,

Query Match	%	Score	2326	DB 11	Length	892	
Best Local Similarity	78.4%	Pred. No. 5.4e-193					
Matches	443	Conservative	10	Mismatches	12	Indels	100
							Gaps
DB	1	TGTANPKYTFMSEIMDAEGR11DIVIDKEHIOPELIFECEVEYIARAGWTPECKYANSI	60				
DB	328	TGTANPKYTFMSEIVDNDAGGI1DIVIDKELVQPELIFECEVEYIARAGWTPECKYANSI	387				
QY	61	LLDRSQRRLQIVLISPELIFEVEDDVMERQRLIESVPSVTPPLIYEETDWINIHIF	120				
DB	388	LLDRSQRRLQIVLISPELIFEVEDDAMDRCRLIESVPSVTPPLIYEETDWINIHIF	447				
QY	121	HVFPQSHHEEIEPIFASCECTGPHLYKTSIILKSKYKRSSGGI.PAPSDRCPIKEET	180				
DB	448	HVFPQSHHEEIEPIFASCECTGPHLYKTSIILKSKYKRSSGGI.PAPSDRCPIKEET	507				
QY	181	ITSGEWEVLGRHGSNIQVDEVRLVYFEGTQDSPLEHLLVYVSVNPEGEVTRLIDRGYSH	240				
DB	508	ITSGEWEVLGRHGSNIWDEARKLIVFEGTQDSPLEHLLVYVSVNPEGEVTRLIDRGYSH	567				
QY	241	SCCISQNHDFPISYXSNQKNHCISLYKLSPEDDPCTCKTEFPATLIDSAGLPDYPP	300				
DB	568	SCCISRHCDPFIKSYXSNQKNHCISLYKLSPEDDPVAHKEFWATLIDSAGLPDYPP	627				
QY	301	EIFPFESTTGTLLYGLMYKPPHDLQPKKYPLVLEIYGGPQ-----	340				
DB	628	EIFPFESTTGTLLYGLMYKPPHDLQPKKYPLVLEIYGGPQVQVQVNNRRKGVYPRNLTLA	687				
QY	341	-----	340				
DB	688	SLGYVVVVVIDRGSCHRGKFEKAGFKYKMGQIEIDDQVGLQYLAQYDFIDLDRVGING	747				
QY	341	-----VAIGAPVTLMIFYDTGTYTERYMGHPDNEQOYLYGSYAM	380				
DB	748	WSYGYLSLMLMORSDFRALVAGAPVTLMIFYDTGTYTERYMGHPDNEQOYLYGSYAM	807				
QY	381	QAEKFPSEPNRLLLHGFUDENVHFAHTSILLSFLVRAKPYDLQIYQOEHSIRVPESG	440				
DB	808	QAEKFPSEPNRLLLHGFUDENVHFAHTSILLSFLVRAKPYDLQIYQOEHSIRVPESG	867				
QY	441	EHVELHLHLHYLQENLGSRIALKYI 465					
DB	868	EHVELHLHLHYLQENLGSRIALKYI 892					
RESULT 7							
ID	Q8WKXD8	PRELIMINARY:	PRT:	863	AA.		
AC	Q8WKXD8:						
DT	01-MAR-2002 (TREMBLrel. 20, Created)						
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)						
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)						
DE	Dipectidyl peptidase 9.						
GN	DP9.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
OX	NCBI_taxid=9606;						
RN	(1)						
RP	SEQUENCE FROM N.A.						
RA	Olsen C., Magtann N.;						
RT	"Identification and characterization of a novel member of the						
RT	dipectidyl peptidase IV-related family.";						
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.						
RN	[2]						

RP SEQUENCE FROM N.A.  
RC TISSUE=skin;  
RA Strauberg R.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF52102; AAL47179.1; -  
DR EMBL; BC037948; AAH37948.1; -  
DR MEROPS; S09.019; -  
DR InterPro; IPR002469; DDPV N term.  
DR InterPro; IPR001375; peptidase\_S9.  
DR InterPro; IPR000379; Ser\_estra\_site.  
DR Pfam; PF00930; DDPV\_N\_term; 1.  
DR Pfam; PF00326; peptidase\_S9; 1.  
SQ SEQUENCE 863 AA; 98263 MW; 40FE0B78E26CED5 CRC64;

Query Match 60.2%; Score 1494.5; DB 4; Length 863;  
Best Local Similarity 50.5%; Pred. No. 8.5e-121;  
Matches 281; Conservative 69; Mismatches 105; Indels 101; Gaps 2;

QY 1 TGTANPKVTFKMEIMDAEGRITIDVDKELIQFELLFEGVEYIARAGMTPEKGYAMSI 60  
DB TGSKNPKIALKLAEPQDSQKIVSTQEKELVQFSSLPFKVEYIARAGMTROCKYAMAM 367  
QY 61 LDRSQTRLOIVLISPELFIPEDDVMERQRLIESVDSVTPLIYEETDIWINHIDIF 120  
DB FLDRPQORLOLVLLPPLFIPESTENEGRLASARAAYPRNVQPVYVEEVTNWMINVIDIF 427  
QY 121 HVPQOSH-EEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFKCIKEEI 179  
DB YPPQAGQGDQFCFLRANECKTGFCFLYRTVLTCTKDQVMTBLSPTBDFKCIKEEV 428  
QY 180 AITSGEWEVLGRGNSIQDVEVRLLVFEETKDSPLEHHLVYVSYNPGVETRLTDGYS 239  
DB ALTSGEWEVLARHGSKIWNBEETLVYFQGTQOTPLHHLVYVSYEAGETVRLTTGFS 488  
QY 240 HSCCISQHCDFPISKSNQKNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYTP 299  
DB HSCSMQSFMPFVSHSVSTPCVAVHYKLSGDDDLHKQPRFMAEMEAACPPDYVP 548  
QY 300 PEIIFSESTGTFLYGLMYKPHDLOPKKXPTVLFYIGSPQ----- 340  
DB PEIIFHTRADVOLYGLMYKPHDLOPKKXPTVLFYIGSPQVOLVNNSPFKIKYLRNLTL 608  
QY 341 ----- 340  
DB ASLGYAVVVIDGRSGCORGLRFGALKNQMGVEIEDQVGLQFVAEKGFIDLSRAVH 668  
QY 341 ----- 340  
DB GMSYGFSLMGLIHKQVFKVAIAGAPVTWMAAYDTGYTERYMDVPENNQGYEAGSVA 728  
QY 380 MOAEKPPSEBNRLLLHGFLENVHFAHTSILSLFVRACKPYDLOIYPOERSIRVPES 439  
DB LHYEKLPNENRLLILHGFLENVHFAHTSILSLFVRACKPYDLOIYPOERSIRCPES 787  
QY 440 GEHYEHLHLHYLOENTL 455  
DB GEHYEVLTLHLFLOEYL 863

RESULT 8

Q8BMT9 PRELIMINARY; PRT; 862 AA.  
AC O8BMT9;  
DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
DE Dipeptidyl peptidase 9 homolog.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=liver;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK050021; BAC34034.1; -  
SQ SEQUENCE 862 AA; 98001 MW; BID566B24A834E8 CRC64;

Query Match 60.0%; Score 1488.5; DB 11; Length 862;  
Best Local Similarity 50.0%; Pred. No. 2.8e-120;  
Matches 278; Conservative 70; Mismatches 107; Indels 101; Gaps 2;

QY 1 TGTANPKVTFKMEIMDAEGRITIDVDKELIQFELLFEGVEYIARAGMTPEKGYAMSI 60  
DB TGSKNPKIALKLAELQDHOQKIVSSCEKELVQFSSLPFKVEYIARAGMTROCKYAMAM 366  
QY 61 LDRSQTRLOIVLISPELFIPEDDVMERQRLIESVDSVTPLIYEETDIWINHIDIF 120  
DB FLDRPQORLOLVLLPPLFIPEVSEKQRAARAAYPRNVQPVYVEEVTNWMINVIDIF 426  
QY 121 HVPQOSH-EEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFKCIKEEI 179  
DB HPPQAGQGDQFCFLRANECKTGFCFLYRTVLTCTKDQVMTBLSPTBDFKCIKEEV 427  
QY 180 AITSGEWEVLGRGNSIQDVEVRLLVFEETKDSPLEHHLVYVSYNPGVETRLTDGYS 239  
DB ALTSGEWEVLSRHSKIMWBEQRLVYFQGTQOTPLHHLVYVSYEAGETVRLTTGFS 487  
QY 240 HSCCISQHCDFPISKSNQKNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYTP 299  
DB HSCSMQSFMPFVSHSVSTPCVAVHYKLSGDDDLHKQPRFMAEMEAACPPDYVP 547  
QY 300 PEIIFSESTGTFLYGLMYKPHDLOPKKXPTVLFYIGSPQ----- 340  
DB PEIIFHTRADVOLYGLMYKPHDLOPKKXPTVLFYIGSPQVOLVNNSPFKIKYLRNLTL 607  
QY 341 ----- 340  
DB ASLGYAVVVIDGRSGCORGLRFGALKNQMGVEIEDQVGLQFVAEKGFIDLSRAVH 667  
QY 341 ----- 340  
DB GMSYGFSLMGLIHKQVFKVAIAGAPVTWMAAYDTGYTERYMDVPENNQGYEAGSVA 727  
QY 380 MOAEKPPSEBNRLLLHGFLENVHFAHTSILSLFVRACKPYDLOIYPOERSIRVPES 439  
DB LHYEKLPNENRLLILHGFLENVHFAHTSILSLFVRACKPYDLOIYPOERSIRCPES 787  
QY 440 GEHYEHLHLHYLOENTL 455  
DB GEHYEVLTLHLFLOEHL 862

RESULT 9

Q8BVG4 PRELIMINARY; PRT; 862 AA.  
AC O8BVG4;  
DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
DE Dipeptidyl peptidase 9 homolog.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=olfactory brain;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium;  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;



```

OY      121 HVFPOSHEE-LEIEPIFASCKCTGFRHLUYITSLTJLKSKCYKRSSGGJ.PAPSDFCPIKEE1 179
Db      220 YFFPQORETTSSAFASAMMARPASAICTYSPPFLKSGCTIDMSPEPSPGDEDFECPIKEE1 279
OY      180 AITSGEMEWLGHGNSNIQYDEVARLVYFEGTKDSPLEHHLVYVSYNNGEVRTLTRGYS 239
Db      280 ALTSGEMEWLARHGSKIMVNEEKLVYFGOTOTPLEHHLVYVSYEAGIYVLTTPGFS 339
OY      240 HSCCTISQHCFFITSKYSNQNPHCVSLYKLSSEDDPTCKTEFWATILDSAGPLDPYTP 299
Db      340 HSCSMSGNQNDMVSHTSSVSTPCCVHVYKLSGDDDBPLHKQRFMSAMMEACCPDPYVP 399
OY      300 PEIESFESTTGTITVGMILKYPHDLQSKKXPTPLTYFYGGQ----- 340
Db      400 PEIFPHHTKSDVRLYGMILYKPHALQPKQKPTVLVFGVGGQVQLVNNSEFKIGYLRNLNTL 459
OY      341 ----- 340
Db      460 ASLGAIVAVVYIDGRGSCQRLRFEGALKNQMGQVEIEDQVEGLQFVAEKYCFIDLSVAIH 519
OY      341 -----VAIAGAPVTLMIFYDTGYTERMYGHPDONGEYVIGSVA 379
Db      520 GWSYGGFLSLMGLIHKPQVFKVAILAGAPVTVMVAAYDTGYTERMYDVBENNQHGEYAGSVA 579
OY      380 MOAEKFPSPERNLLLLHGFLENVHFANHSILSLFVLRACKPYPDQIY 427
Db      580 LHVEKLPNEPNRLLILHGFLENVHFHNHFNFLSOLIRGKQPQOJEBY 627

```

RESULT	ID	PRELIMINARY;	PRT;	508 AA.
075273	075273			
AC	075273;			
DT	01-NOV-1998	(TRENMBLrel. 08, Created)		
DT	01-NOV-1998	(TRENMBLrel. 08, Last sequence update)		
DT	01-MAR-2003	(TRENMBLrel. 23, Last annotation update)		
DE	R26984_1 (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,			
RA	Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwogen S.,			
RA	Pfan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,			
RA	Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,			
RA	Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,			
RA	Coeffield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,			
RA	Krommuller B., Atellano A., Montgomery M., Ow D., Nolan M., Trong S.,			
RA	Kobayashi A., Olsen A.S., Carraro A.V.;			
RT	"Sequence analysis of a 2.5 Mb region in 19p13.3.";			
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC005594; AAC33801.1; -.			
DR	MEBOPS; 509.019; -.			
DR	InterPro; IPR000379; Ser_centre_site.			
FT	NON_TER			
FT	1			
SEQ	SEQUENCE	508 AA;	57750 MW;	2P4D6645BE2D2C89 CRC64;

Query Match	47.8%;	Score 1187.5;	DB 4;	Length 508;
Best Local Similarity	44.4%;	Pred. No. 1.8e-94;		
Matches 241;	Conservative 64;	Mismatches 93;	Indels 145;	Gaps 7;

```

Qy 23 I I D V I D K E I O F E L L I B E C V E Y I A R G W T P E G K V A M S I L D S Q R T L O V I L S P E I F V 82
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 I V S T O K E L V O F S S L F P K V E Y I A R G ----- A M A M L D R Q O M L O V L P L P A L F I S 53
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 83 E D D W M E R Q R L I S V P D S T P L I I Y E T T I M I N I D I F H A V P O S H - E E I E I P F A S E C K T 141
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 54 T E N E Q R L A S A R A V P R N O V P Y V E V I T W M I N V H D I P Y P F O S E E D E D I C F R A N E C K T 113
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 142 G F R H L Y K T T S I L K E S K Y R K S S G L P A P S D F C P K I E I L I T S G E V E L R N H G S N I O V D E V 201

```

```

Db      114 GCCHLYKTTAVLKSGCYDMKSEPFSG-----EGEGSLTNAIW-----VNEE 154
Qy      202 RRLVYFEGTKOSPLEHNLVVSYYVNPGVTRLTDRGYSHSCCISOHCDFEISKYSNQXP 261
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
155 TGLVYFQGTQDTRLEHNLVVSYYEAAGEIVRLTTPGFSSCSMSQNFDMVSHYSSVSTR 214
Qy      262 HCVSLYKLISSPEDDPTCKTKEFWAATILDSAGRLPDYTPPELISFESBTGFTLYGMLYKPH 321
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
215 PCVHYTKLISGPPDDDLHKQPRFWMAMMEA-----KIFHFHTRSDVRLYGMJYKPH 265
Qy      322 DLQPGKKYPTVLFITYGPP-----340
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      266 ALQPKKHPVLFVYGGPQVQLVNNSEFKIKYLRLNTLASLGVAVVYIDGRSCQQRGRF 325
Qy      341 -----Y 341
Db      326 EGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRAVHGMSTYGGFLSLMGLIHKKPQVFKV 385
Qy      342 AIAQAPVLTMTFYDPTGYTERWYMGHPDQNEQGYVLSVAMQAEKFPSEPNRLLLHGFIDE 401
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      386 AIAQAPVTVMAAYDYDGYTERVYDVDPENNQHGYEAGSVLAHVEKLPREPNRLLLHGFIDE 445
Qy      402 NVHFARTSLILSLVRAGKPYDL-----QIYQENHSIRVPSGGHYELHLHLYIQ 452
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      446 NVHFHTNFWLSQLIRAGKPYQLQVALPVSQIYPNEHSIRCPSSGGHYEVLTLHFIQ 505
Qy      453 ENL 455
      :|
Db      506 EYL 508

```

<b>RESULT 13</b>					
ID	Q9BVR3	PRELIMINARY:	PRT:	439 AA.	
AC	Q9BVR3;				
DT	01-JUN-2001 (TrEMBLrel. 17, Created)				
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)				
DI	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)				
DE	Hypothetical protein (Fragment).				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_TaxId=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta;				
RA	Strausberg R.;				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC000970; AAH00970.1; ..				
DR	MEROPS; S09.019; ..				
DR	GeneW; HGNC:18648; DPP9.				
DR	InterPro; IPR001375; Peptidase_S9.				
DR	InterPro; IPR000379; Set_eutr_site.				
DR	Pfam; PF00136; Peptidease_S9; 1.				
KW	Hypothetical protein.				
FT	NON_TER				
SO	SEQUENCE_439_AA; 49926_MW; A18BA9EI2092BAF_CRC64;				

Query Match 46.5%; Score 1154.5; DB 4; Length 439;  
Best Local Similarity 50.6%; Pred. NO. 1.1e-91;  
Matches 222; Conservative 44; Mismatches 72; Indels 101; Gaps 2;

```

QY 118 IIFHFPPSH-EEIEFIFPASECKGFHLYITSLIKESKTKRSGCLPANSDFPCPLK 176
Db 1 DIFHFPPSBEGBDELCFLRANECKTGFCFLHYVNLVNLGQYDWSPEFSGEDEFCKPLK 60
QY 177 EEIAITSGEWEVLGRHGSNIQVDEVRLRVYFEGTDSPLEHLLVYVSYNPGSVRTLDR 236
Db 61 EEIAITSGEWEVLARHGSKIWNBEETKLVFQGTOTPLEHLLVYVSYEADGIVRLTTP 120
QY 237 GYHSOCCISQHCDFIISKYSNOKNPHCSLYKLTSPDEDDPTCTKTEFMAATILDSAGPLTD 296

```

Db 121 GFSHSCSMSONFPMFVSHYSSVSTPPCVHYKLSGPDDBLHKQPRFWMASMEASCPD 180  
Qy 297 YTPPEIFSFESTGFTLYGMLYKPHDLOPKKYPVTLFIYGPQ----- 340  
Db 181 YVPEIFPHFTRSDVRLYKMYKPHALOQPKKHFTVLFYGGQVQVLNNSPFKIKYLR 240  
Qy 341 ----- 340  
Db 241 NTLASLGAVVVIDRGSCORGLRFEGALKNOMQVEIEDQVEGLQFVAEKYFIDLRSV 300  
Qy 341 -----VAIAGAPVTLMTIYFDGYTFRYVNGHGDQNEQYTLG 376  
Db 301 AIHGWSYGLSLMGLIHKQVPEKVALAGAPVTVMAVDTGYTERYMDVENNOQHYEAG 360  
Qy 377 SVAMQAEKFPSEPNRLLLHGLFDENVHFAHTSILSLFVRAGKPYDLQIYQBRHSIRV 436  
Db 361 SVALLHVEKLPNERRLILHGLFDENVHFAHTSILSLFVRAGKPYDLQIYQBRHSIRV 420  
Qy 437 PESGEHYELHLHYLOENL 455  
Db 421 PESGEHYEVTLLHFLQEVYL 439

## RESULT 14

Q9HBM2 PRELIMINARY; PRT; 360 AA.  
AC O9HBM2;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Dipeptidyl peptidase 8 (Fragment).  
GN DPP8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20467194; PubMed=11012666;  
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,  
RA Gorrell M.D.;  
RT "Cloning, expression and chromosomal localization of a novel human  
RT dipeptidyl peptidase (DPP) IV homolog, DPP8".  
RL Eur. J. Biochem. 267:6140-6150 (2000).  
DR EMBL; AF221637; AAC29769.1; -  
DR MEROPS; S09.018; -  
DR InterPro; IPR001375; Peptidase\_S9.  
DR InterPro; IPR001379; Ser\_estrs\_site.  
DR Pfam; PF00326; Peptidase\_S9; 1.  
FT NON\_TER 1  
FT TER 360  
SQ SEQUENCE 360 AA; 41070 MW; CF81COBB61423E22 CRC64;

Query Match 34.7%; Score 860.5; DB 4; Length 360;  
Best Local Similarity 40.5%; Pred. No. 2.8e-66;  
Matches 186; Conservative 0; Mismatches 0; Indels 273; Gaps 2;  
Qy 1 TGTANPKVTFKMEIMIDAGRIIVDKELIQFELLFEGVEYIARAGWTPGKYAWSI 60  
Db 75 TGTANPKVTFKMEIMIDAGRIIVDKELIQFELLFEGVEYIARAGWTPGKYAWSI 60  
Qy 61 LLDRSQTRLOIVLISBELFIPVEDDVMERQRLIESVPDSVPLIYEETTDININHDIF 120  
Db 99 ----- 98  
Qy 121 HVEPQSHHEIEIFIFASECKTGRHLKYKTSILKSKYKRSGLPAPSPDKPIKEEIA 180  
Db 99 ----- 98  
Qy 181 ITSGEWEVLGRHSGNIQVDEVRRLVYFEGTKDSPLEHNLVYVYVNGEVTRLTRDGYSH 240  
Db 99 -----VDEVRRLVYFEGTKDSPLEHNLVYVYVNGEVTRLTRDGYSH 141

Qy 241 SCGISQHCDFEISKYSNKNXPHCVSLYKLSPPEDDPTCKTKFMTATLDSAGLPDYTPP 300  
Db 142 SCGISQHCDFEISKYSNKNXPHCVSLYKLSPPEDDPTCKTKFMTATLDSAGLPDYTPP 201  
Qy 301 EIFSPESTGFTLYGMLYKPHDLOPKKYPVTLFIYGPQ----- 340  
Db 202 EIFSPESTGFTLYGMLYKPHDLOPKKYPVTLFIYGPQVQVLNNSPFKIKYLR 261  
Qy 341 ----- 340  
Db 262 SLGTVVVVIDNRGSCORGLRFEGALKNOMQVEIEDQVEGLQFVAEKYFIDLRSV 321  
Qy 341 -----VAIAGAPVTLMTIYFDGYTFRYVNGHGDQNEQYTLG 376  
Db 322 WSYGYGLSLMGLIHKQVPEKVALAGAPVTVMAVDTGYTERYMDVENNOQHYEAG 360

## RESULT 15

Q9VC20 PRELIMINARY; PRT; 1042 AA.  
AC Q9VC20;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE CG3744 protein.  
GN CG3744.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chame M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,  
RA Borokov D., Botchan M.R., Bouck J., Brostein P., Brotier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fessler C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
RA Jalali M., Kalish F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merklou G., Miliushina N.V., Mobarry C., Morris J., Mosherfi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
RA Palatcollo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RT Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RL "The genome sequence of Drosophila melanogaster.";  
Science 287:2185-2195 (2000).

DR EMBL: AE003749; AAF56357.1; -.  
DR MEROPS: S09.016; -.  
DR FlyBase: FBgn0039240; CG3744.  
DR InterPro: IPR002469; DPIP\_N\_term.  
DR InterPro: IPR001375; Peptidase\_S9.  
DR InterPro: IPR000379; Ser\_estrs\_site.  
DR Pfam: PF00930; DPIP\_N\_term; 1.  
DR Pfam: PF00326; Peptidase\_S9; 1.  
DR Pfam: PF00326; Peptidase\_S9; 1.  
SQ SEQUENCE 1042 AA; 116706 MW; 5B915D5C365DP937 CRC64;

Query Match 34.0%; Score 845; DB 5; Length 1042;  
Best Local Similarity 34.4%; Pred. No. 2.8e-64;  
Matches 201; Conservative 76; Mismatches 155; Indels 152; Gaps 15;

QY 1 TGTANPKVTFKMEIIMDAAGRIIDVIDDELIQPEIL--FEGVEYIARAGWTFEGKTAW 58  
DB 477 TGSFNKSKLKLQVQFVNLALQVSEIATKDL--PYSLLAVSWLEYIVRGWTFDAKYYW 534  
QY 59 SILLDRSQTRLQIVLISPELFIPEDDVMERQRLIESVPD-----SVTPL-IYEE 108  
DB 535 VQGLDRKQQLDYLILPLDNF--CEGSSQSVSTPTDSIGHSWRSLSYRTITPLQVIYTE 592  
QY 109 TTDIWINIHIDIFHVPOSHDEEIEFIASECKTGFRHLKYITSILKSKYKRSSGGLPAP 168  
DB 593 RSDSWINVDMLH-FIDLLETSTVFLWASE-ETGFRHLVLTASLLSQ---ANGQDP 646  
QY 169 -----SDFKPIKEELAITSGEWEVLGRHSNIQVDEVRILYFEETKDSPL 216  
DB 647 GSVGAQPSFVDSALQPIRLINKKALTSGEWEVLAR--NLWDKANKLVYFVGLRDTPL 703  
QY 217 HHLVYVSVNPGEVTRLTDRGYSHSCISQCHDFISKYSN--QKNPHCVSLYKLSPEDD 275  
DB 704 KHLVYVSLRPERHRIILTERGYSLVEFPDQCKMLLVYCNIGRLPSCKM-----RVN 757  
QY 276 PTCKTEKEFWATTIDSAG-----PLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKX 328  
DB 758 QTCNMGVANGIQISLVGYLHEGKPEPOYC-POIFSPOLPSGDIYVAMVFKPHNFELGYK 816  
QY 329 YPTVLFIYGGPQ----- 340  
DB 817 YPTVLNVVGGPEVQVTNNNTFKGHQLRMHMLAAQYCVICIDSRGSRHGRKFESHIRGR 876  
QY 341 -----VAIAGAPV 348  
DB 877 MGQVELTDQVDALRSLSQGLGYIDMDRVAIHGWSYGYLSLMLGVQYPKIFKAIAGAPV 936  
QY 349 TLMIFYDTGYTERYMGHPDQNEQGYLYGSVAMQAEKFPSEPNRLLLHGFLENTVPAHT 408  
DB 937 TMEVYDTGYTERYMDMPONNEAGSYLLEVNSFPEDKRLLLHGLIDENVHFCHT 996  
QY 409 SLLSFLVRAKRPYDLOIYPOERHSTRVPESGHEVELHLHYLO 452  
DB 997 SRLISALNANKPEYVHLFPERHSLNLESNKNYETKLISFLQ 1040

Search completed: October 15, 2003, 17:13:45  
Job time : 31.7397 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:15:46 (Search time 5076.73 seconds  
(without alignments)  
3747.092 Million cell updates/sec

Title: US-10-070-464-5  
Sequence: 1 TGTATPKVTFKMSIWMIDAE.....HLHLHYQENIGSRIALKVI 465  
Perfect score: 2482

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO/US10070464/runat.15102003.113553.24810/app.query.fasta\_1.2652  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ctc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10070464 @CCN 1.1 13561 @runat.15102003.113553.24810 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEUDERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONELOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_scs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_va: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_scs: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_hlg\_hum: \*  
31: em\_hlg\_inv: \*  
32: em\_hlg\_other: \*  
33: em\_hlg\_mus: \*  
34: em\_hlg\_pin: \*  
35: em\_hlg\_rod: \*  
36: em\_hlg\_mam: \*  
37: em\_hlg\_vrt: \*  
38: em\_sy: \*  
39: em\_hlgo\_hum: \*  
40: em\_hlgo\_mus: \*  
41: em\_hlgo\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2482	100.0	1669	9 AF221636	AF221636 Homo sapi
2	2482	100.0	2668	6 AX405771	AX405771 Sequence
3	2482	100.0	2842	6 AX405770	AX405770 Sequence
4	2455	98.9	2778	9 AK000280	AK000280 Homo sapi
5	2447.5	98.6	3130	9 BC030688	BC030688 Homo sapi
6	2446.5	98.6	4523	6 AX608731	AX608731 Sequence
7	2441.5	98.4	2161	6 BD157001	BD157001 Primer fo
8	2441.5	98.4	2161	9 AK027826	AK027826 Homo sapi
9	2432	97.6	2649	6 AX354795	AX354795 Sequence
10	2432	97.6	2656	9 AY172659	AY172659 Sequence
11	2422	97.6	2671	6 AX608725	AX608725 Sequence
12	2422	97.6	3106	6 AX342633	AX342633 Sequence
13	2422	97.6	3127	9 AF221634	AF221634 Homo sapi
14	2422	97.6	3143	6 AX354793	AX354793 Sequence
15	2422	97.6	4335	9 BC040203	BC040203 Homo sapi
16	2418.5	97.4	4576	6 AX608743	AX608743 Sequence
17	2402	96.8	4829	6 AX608735	AX608735 Sequence
18	2362	93.7	4799	10 BC043124	BC043124 Mus muscu
19	2244.5	90.4	2510	6 AX338497	AX338497 Sequence
20	2149	86.6	4685	6 AX608745	AX608745 Sequence
21	1494.5	60.2	2602	9 AY172660	AY172660 Homo sapi
22	1494.5	60.2	2617	6 AX608727	AX608727 Sequence
23	1494.5	60.2	3000	9 AF542510	AF542510 Homo sapi
24	1494.5	60.2	3716	6 AX480934	AX480934 Sequence
25	1494.5	60.2	4219	6 AX608751	AX608751 Sequence
26	1494.5	60.2	4232	9 BC037948	BC037948 Homo sapi
27	1494.5	60.2	4295	9 AF452102	AF452102 Homo sapi
28	1494.5	60.2	4302	6 AX608747	AX608747 Sequence
29	1491.5	60.1	2676	6 AX524928	AX524928 Sequence
30	1445	58.2	4180	6 AX608759	AX608759 Sequence
31	1445	58.2	4263	6 AX608757	AX608757 Sequence
32	1386	55.8	4309	6 AX608737	AX608737 Sequence
33	1377.5	55.5	1197	9 AF221635	AF221635 Homo sapi
34	1342.5	54.1	4076	6 AX608755	AX608755 Sequence
35	1342.5	54.1	4159	6 AX608753	AX608753 Sequence
36	1338	53.9	3243	9 HSM805448	AL834376 Homo sapi
37	1293	52.1	4037	6 AX608763	AX608763 Sequence
38	1293	52.1	4120	6 AX608761	AX608761 Sequence
39	1274.5	51.3	2261	6 AX713363	AX713363 Sequence
40	1274.5	51.3	2261	9 AK054656	AK054656 Homo sapi
41	1154.5	46.5	2546	9 BC000970	BC000970 Homo sapi
42	1055	42.5	753	6 AX524935	AX524935 Sequence
43	860.5	34.7	1083	9 AF221637	AF221637 Homo sapi
44	841.5	33.9	3740	3 BT001499	BT001499 Drosophila
45	673.5	27.1	215711	3 AC008205	AC008205 Drosophila

RESULT 1

## ALIGNMENTS

AF221636  
 LOCUS AF221636 1669 bp mRNA linear PRI 05-NOV-2000  
 DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, partial cds,  
 alternatively spliced.  
 ACCESSION AF221636  
 VERSION AF221636.1 GI:11095191  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1669)  
 AUTHORS Abbott,C.A., Yu,D.M., Woollett,E., Sutherland,G.R., McCaughan,G.W.,  
 and Gorrell,M.D.  
 TITLE Cloning, expression and chromosomal localization of a novel human  
 dipeptidyl peptidase (DPP) IV homolog, DPP8  
 JOURNAL Eur. J. Biochem. 267 (20), 6140-6150 (2000)  
 MEDLINE 20467194  
 PUBMED 11012666  
 REFERENCE 2 (bases 1 to 1669)  
 AUTHORS Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver  
 Centre, Centenary Institute of Cell Biology and Cancer Medicine,  
 Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia  
 LOCATION/Qualifiers  
 1. 1669  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /map="15q22"  
 /chromosome="15"  
 /map="15q22"  
 /clone="T8"  
 /tissue type="placenta"  
 /clone\_lib="lambda triplex tm: Clontech"  
 <1. 1669  
 /gene="DPP8"  
 <1. 1399  
 /gene="DPP8"  
 /note="serine protease; member of prolyl oligopeptidase S9b  
 family belonging to enzyme clan SC; isoform missing the  
 catalytic serine required for enzyme activity; additional  
 splice variants of this mRNA are found in Genbank  
 Accession Numbers AF221634, AF221635 and AF221637"  
 /codon start=2  
 /product="dipeptidyl peptidase 8"  
 /protein\_id="AAG29768.1"  
 /db\_xref="GI:11095192"  
 /translation="TGTANPKQTFKMSIMIDAEGRITDIVIDKELIQPELLPEGVVEY  
 IARAGTPEGKYAMSLIDRSQRLQIVLISPELTPVEDDVMERQLIESVDSVTP  
 LRIYETTDIWINIHDIHFVFPQSHHEEIEPIFASCKTGFPHLYITSLKESYKR  
 SSGGAPSPDFKPIKEEIAITSGEWEVNSIQOVSRRLVFEETKDSPEHH  
 LYVGVNPGVEVRLTRIDRGYSKCCISQHCDFISKYSQKPKHCISLYKLSPEDDP  
 TCKTKFNAITIIDSGRLPBDYTPREIFSESTGTFLYGMLYKPHDLOQCKPYTLF  
 IYGGPOVAIAGAPVTDIMTYDGTERTYGHFPONEQGYLDSVAQAQKPFSEPRLL  
 LLHGFLDENVHFATSLILSLFLVRAGKPYDLQIYQERHSTRVESGEHYELHLHY  
 LOENLGSRIALAKVI"  
 BASE COUNT 524 a 336 c 336 g 473 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,466-236 Length: 1669  
 Score: 2482.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-070-464-5 (1-465) x AF221636 (1-1669)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPhelyMetSerGluIleMetIleAspAlaGlu 20  
 DB 2 ACAAGTACAGCAATCTTAAGTCACTTTTAAGATGTCAGAAATATATGATTGATGCTGA 61

QY 21 GIYArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleuPheGlu 40  
 DB 62 GGAAGATCATAGATGTCATAGATAAGAACTATTCACCTTTGAGATTCTATTGAA 121  
 QY 41 GIYValGluTyrIleIleAargAlaGlyTyrThrProGluGlyLysTyrAlaTrpSerIle 60  
 DB 122 GGAGTTGAATATATTGTCAGAGCTGATGATGATCTCCAGAGAAATATATCTGCTCANC 181  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGluIleValIleuIleSerProGluLeuPheIle 80  
 DB 182 CTACTAGATCGCTCCAGACTGCTGCTACGAAAGTGTGATCTCACCTGAATTAATTATC 241  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 242 CAGATGAAAGATGATGTTTATGAAAGGAGAGAACTCATGATGATGATGCTGATCTG 301  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 302 ACGCCACTAATTATCTATGAGAAACAACAGACATCTGGATTAATTCATGACATCTTT 361  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleSerGluCysLys 140  
 DB 362 CATGTTTTCCCAAGTCACGAAAGAAATTGAGTTATTTTGGCTTGTAATGCAA 421  
 QY 141 ThrGlyPheArgHisLeuTyrIleTyrHisIleThrSerIleLeuLeuGluSerIleTyrLysArg 160  
 DB 422 ACAAGTTCCTCCATTTATACAAATATCATCTATTTTAAAGAAAGCAATATTAACGA 481  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluIleAla 180  
 DB 482 TCACAGTGGGCTGCTGCTCCCAAGTATTCAGTGTCTTATCAAAAGAGATAGCA 541  
 QY 181 IleThrSerGlyGluThrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 DB 542 ATTACAGTGGGAAATGGAGATCTTGCGCGCAGTGAATTAATTCAGATGATGAA 601  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyr 220  
 DB 602 GTCAGAGGCTGTATATTATTAAGGCACCAAGACTCCCTTTAGAGATCACCTGTAC 661  
 QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 662 GTAGTCAGTTACGTAAATCTCTGAGAGGTCGACGAGCTGACCTGCTACTACAT 721  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 DB 722 TCTTGCTGATCAGTCAGACGTGACCTTCTTATTAAGTAAATATGTAACCAAGAAAT 781  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 782 CCACACTGTGTCTCCCTTACAAAGATCACTCACTCGAAGATGACCCAACTTGCAAAACA 841  
 QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 DB 842 AAGGAATTTGGGCCACCACTTTTGATTCAGAGGCTCCCTTCCTGACATATCTCTCCA 901  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 902 GAAATTTCTCTTTGAAAGTACTAGATTTAATGATGAGATGCTCTTACAAAGCT 961  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 962 CATGATCTACAGCTCGAAAGAAATATCTACTGTCGCTGTTCATATATGCTGCTCAG 1021  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrPheMetYAspThrGlyTyrThrGlu 360  
 DB 1022 GTTGCTATTGCTGGGACCCCACTCACTCTGAGATCTTATGATACAGAGATACAGGAA 1081  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 1082 CATTATATGGGTCACCTGACCAAGATGAACAGGCTATTACTTATGATCTGTGGCCATG 1141



QY 381 GlnAlaGluPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
DB 1142 CAAGCAGAAAAGTCCCTCTGACCAATCGTTACTGCTCTTACATGCTTCCGAA 1201  
QY 401 GlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
DB 1202 GAGAAATGTCATTTTSCATACATACAGATATATTACTAGATTTTACTGAGGGCTGGAAG 1261  
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
DB 1262 CCATATGATTTACAGATCTTACTCTCAGAGAGACAGACATAGAGACTTCTGATGAGGA 1321  
QY 441 GlnHisTyrGluLeuHisIleLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
DB 1322 GAACATTAATGAACTGATCTTTTGACATACCTTCAAGAAAACCTTGATCAGCTATTGCT 1381  
QY 461 AlaLeuLysValIle 465  
DB 1382 GCTCTAAAGTGATA 1396  
RESULT 2  
AX405771 2668 bp DNA linear PAT 14-JUN-2002  
LOCUS Sequence 186 from Patent WO0222660.  
DEFINITION AX405771  
ACCESSION AX405771  
VERSION AX405771.1 GI:21438981  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,  
Xue, A.D., Yang, Y., Weinman, T., and Drmanac, R.T.  
TITLE Novel nucleic acids and polypeptides  
JOURNAL Patent: WO 0222660-A 186 21-MAR-2002;  
HYSEQ, INC. (US)  
FEATURES  
source location/Qualifiers  
1..2668  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
234..2408  
/codon\_start=1  
/protein\_id="CAD34878.1"  
/db\_xref="GI:21438982"  
/translation="MAAAMEEQQLGVEIFETADCEENTESODRPKLEPFYVERYSMQ  
LKLKLADTRKHYGMKAKAPHFMEYKRDPGPSDRILYILAMGENSENTLIFYSEI  
PKTIRAAVLM,SMKPLDL,FOOOLPRPL,VERTSCPNTRMDPKLCPADMDWIAFIHNS  
DIWISNITVREERRLTYVNEELANMEEDRSAGVATFVIOEFDYSGWMPKATTT  
DSGKITRLIYLENESEVEITIHVTSPLMETRADSFYIKGTANPKVTFPMSEIMI  
DAEGRIIDVIDELIQPEILLEGVEYLARAWTPGKAWMSILLDRSGTRQLIVLIS  
PELFIPEVDVVERORLIESVDSVPLIYEETDIWNIHIDIFHVFQSHBEIER  
IFASEKTFGRLLYKITSILKESKYKRSGLPARSDFCPIKEELAIKSGMEVILGR  
HGSNIQVDEVRLLYVEGTRKDSPLLEHLVVSVANGVETRLIDGSHSCCISIOHD  
FRISKYSNOKNHCUSLYKLSPEDDPTKTKPEWATITLDSAGPLPDTYPTPTSPES  
TTGFTLGLMLYKPHDQPKKYPTVLTFGQVATAGAPVLMIFDYGYETRWGH  
PDNQGTYLGSVAAQEKPRSEPNRLHLHGFIDENVHFAHTSLISFLVAGKPYD  
LQIYPOERSHSIVPESGEHYEHLHLHYLOENIGSRIALKVI"

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
DB 1011 AONGRTACGACAAATCTTAAGTCACTTTAAAGATCTCAGAAATATATGATGCTGAA 1070  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
DB 1071 GGAAGGATCATGATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1130  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysValIleAspSerIle 60  
DB 1131 GAGATGGAATTAATATGCGAGACGATGATGATGATGATGATGATGATGATGATGAT 1190  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1191 CTACTAGATGCTCCAGACGCTGCTTACAGATGATGATGATGATGATGATGATGATGAT 1250  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1251 CCAGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1310  
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
DB 1311 AGCCACTAATTATCTATGAGAAACACACACATCTGATTAATATTCATGACATCTTT 1370  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1371 CATGTTTTTCCCAAGTACAGACAGAGAAATGAGTTATTTTTCCTGATATCCAA 1430  
QY 141 ThrGlyPheArgHisIleuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
DB 1431 ACAGGTTTCGTCATTATTAACAAATTAATGATGATGATGATGATGATGATGATGATGAT 1490  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
DB 1491 TCCAGTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1550  
QY 181 ILeThrSerGlyGluTrogIuValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1551 ATTTACAGTGGGATGGAGATGTTCTGGCCGAGATGATGATGATGATGATGATGATGAT 1610  
QY 201 ValArgArgLeuValTyrPheGluGlyLysThrLysAspSerProLeuGluHisIleLys 220  
DB 1611 GTGAGAGGCTGTATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACCTGTAC 1670  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
DB 1671 GTAGTCAGTTACGTAAATCTGAGAGGTGACAAAGCTGACCGTGTACTACAT 1730  
QY 241 SerCysCysIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGlnLysAsn 260  
DB 1731 TCTTGCTGATCAGTACGACATGCTGATCTTTATTAAGTATGATGATGATGATGATGAT 1790  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1791 CCACATGCTGTCTCTTTACAGAGTATCAATCTGAGAGAGACCAACTTGCAAAACA 1850  
QY 281 LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
DB 1851 AAGGATTTTGGGCACCACTTTGATTCAGACGCTCTCTTCTGACTATACTCTCA 1910  
QY 301 GlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 1911 GAAATTTCTCTTTGAAAGTACTAGATTTTATGATGATGATGATGATGATGATGATGAT 1970  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
DB 1971 CATGATCTACAGCCCGAAAGAAATATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlu 360  
DB 2031 GTTGCATTTGCTGGGGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2090  
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGluGlyTyrTyrLeuGlySerValAlaMet 380

DB 2091 CGTTATATGCGTACCCCTGACCGAATGAAACAGGCTATTACTAGATCTGGCCATG 2150  
QY 381 GlnAlaGluysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
Db 2151 CAGCGAAGAAAGTTCCCTCTGTAACCAATCGTTACTGCTTACACAGGTTTCCGAGAT 2210  
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
Db 2211 GAGAAATGTCATTTTCCACATACAGATATATTACTGAGTTTCTTAGTGAGGCGCTGAAAG 2270  
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
Db 2271 CCATATGATATTACAGATCTATCTCTCAGAGAGACACAGCATAGAGCTTCTGATCGGGA 2330  
QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
Db 2331 GAACATTAATGAACTGCATCTTTTGCACTACCTTCAAGAAACCTTGAGATCAGTATTGCT 2390  
QY 461 AlaLeuLysValIle 465  
Db 2391 GCTCTAAAGTGATA 2405

## RESULT 3

AX405770 LOCUS AX405770 2842 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 185 from Patent WO0222660.  
ACCESSION AX405770  
VERSION AX405770.1 GI:21438979  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,  
Xue, A.U., Yang, Y., Wehrman, T. and Dymnac, R.T.  
TITLE Novel nucleic acids and polypeptides  
JOURNAL Patent: WO 0222660-A 185 21-MAR-2002;  
HYSEQ, INC. (US)

FEATURES  
source

1..2842 Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
234..2582  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD34877.1"  
/db\_xref="GI:21438980"

BASE COUNT 857 a 592 c 635 g 758 t  
ORIGIN  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

Alignment Scores:  
Pred. No.: 3.03e-236 Length: 2842  
Score: 2482.00 Matches: 465  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0  
US-10-070-464-5 (1-465) x AX405770 (1-2842)  
QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
Db 1185 ACGAGTACAGCAAACTCTTAAGCTCACTTTAAGATCTCAGAAATATAGTATGCTCGAA 1244  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
Db 1245 GGAAGGATCATAGATGCTATAGATAGGAACATTAATCACTTTGAGATTCATTATGAA 1304  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
Db 1305 GAGGTTGATATATTATGCGAGACTGGATGATCTCCGAGAGAAATATCTTGCTCATC 1364  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1365 CTAAGTATGATGCTCCAGACTGCTGCTACAGATAGTGTGATCTCACTGAATTAATTATC 1424  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
Db 1425 CCAAGTACAGATGATGATTTGGAAGGACAGACATCATTTGATGCTGCTGATCTTG 1484  
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
Db 1485 ACGCCACTATATTATCTATGAAAGAACACAGACATCTGGATTAATATCATGATCATCTTT 1544  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
Db 1545 CATGTTTTCCCAAGTACAGAGAGAAATGAGTTATTTTGGCTTGATGATCAAA 1604  
QY 141 ThrGlyPheArgHisLeuTyrIleTyrSerIleLeuLysGluSerIleTyrLysArg 160  
Db 1605 ACGGTTTCCGATCATTTATACAAATTTACATCTATTTTAAAGAAACAAATTAACAA 1664  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
Db 1665 TCCAGTGGGAGGCTGCTGCTGCTCAAGATGATTTCAAGTGTCTTCAAGAGAGATGAGA 1724  
QY 181 IleThrSerGlyGluThrGluValLeuGluValArgHisGlySerAsnIleGlnValAspGlu 200  
Db 1725 ATTACCAAGTGGATGATGGAGATTTCTTGCCGCGATGATCTTAATATCCAAAGTATGAA 1784  
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLysLeuTyr 220  
Db 1785 GTCAGAGGCTGATATATTATTAAGCACCAAGACTCCCTTTAGAGATTAACCTGTAC 1844  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
Db 1845 GTAGTCAGTTACGTAATCTCTGAGAGAGTGAACAAGCTGACCGTGGCTATCTACAT 1904  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnIleLysAsn 260  
Db 1905 TCTTGCTGATAGTACAGACTGTACTTCTTATTAAGTAATTAAGTAACAGAGAAAT 1964  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
Db 1965 CCACACTGCTGCTCCCTTTACAGACTATCAAGCTCTGTAAGATGACCAACCTGCAAAACA 2024  
QY 281 LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
Db 2025 AAGGATTTTGGCCACACATTTTGATTCAGACAGCTCTCTTCACTATATCTCTCCA 2084  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
Db 2085 GAAATTTTCTCTTTGAAAGTACTGATTTTACATGTAATGGAAGTCTTACAGAGCT 2144  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
Db 2145 CATGATCTACAGCTCGAAGAAATATCTCACTGCTGCTGATATATGATGCTGCTCAG 2204  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlu 360

Db 2205 GTTCTCTATGCTGGGGCCCGCATCTCTGTGATCTTCTATGATACAGATACACGGAA 2264  
Qy 361 ArGTyTetGtYHtSPRoAsPGLnaSnglUnglYtYrYrLeuGlYserValAlMet 380  
Db 2265 CGTATATGGGTACCCCTGACAGATGAACAGGCTATTACTTAAGGATCTGTGCGCATG 2334  
Qy 381 GlnAlaGlUlyrPheProSerGluProAsnArgrLeuLeuLeuLeuHtSglYpHeLeuAsp 400  
Db 2325 CAACGAGAAAGTTCCCTCTGAAACCAATGCTTACGCTTACATAGCTTCTCGAT 2384  
Qy 401 GlnAsnValHisPheAlaHisThrSerlleLeuLeuSerPheLeuValArgAlaGlylys 420  
Db 2385 GAGATGTCCTCATTTGGACATACAGATATATTACTGATTTTAAAGTGAAGGCTGAAAG 2444  
Qy 421 ProTyraSPLeuGlnlleYrProGlnGlnArgrHisSerlleArgValProGlnSerGly 440  
Db 2445 CCATATGATTTTACAGATCTATCTCAGAGAGACACAGCATTAAGATTCCTGAATCCGGA 2504  
Qy 441 GlnHisTyrglUleuHtSleuLeuHtSlyrLeuGlnGlnLeuLeuGlySerArglleAla 460  
Db 2505 GAACATTATGACCGCATCTTTTGCACACTTCAAGAAACCTTGATCATCGTATTGCT 2564  
Qy 461 AlaLeuLySVallle 465  
Db 2565 GCTCTAAAGTGATA 2579  
RESULT 4  
AK000290 2778 bp mRNA linear PRI 22-FEB-2000  
LOCUS Homo sapiens cDNA FLJ20283 f1s, clone HEP04088.  
DEFINITION AK000290.1 GI:7020272  
ACCESSION AK000290.1 GI:7020272  
VERSION c1igo capping; f1s (full insert sequence).  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2778)  
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.  
Direct Submission  
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).  
FEATURES  
Source 1. .2778  
/location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP04088"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_lib="HBP"  
/note="cloning vector pME18SFJ3"  
163. .1572  
/note="unnamed protein product"  
/codon\_start=1  
CDS

/protein\_id="BA091059.1"  
/db\_xref="GI:7020273"  
/translation="MAAMETEOLGVEIFPTADCEENIESODRPEKLEPFYVERYSWSO  
LKULIATRTKHGMAKAPHDPIFVRANDPDGHSRIITYLANSGERENTLYRSEI  
KPTINRAVLMSKPLDLFOATLDGWTMSREBELRERKICTGVIGASTDYHQS  
TFLEQASGSIYHVNDGGPQFTQOPLAPNIVETSPVIRMDPKCPADPMIAFHSN  
DIWISNIVTREERLTYVHNELANMEDASAGVATVLOEFPDYGVMCPAETT  
PSGKILRIIYEENDESEVEIHYTSPMLERRRDSRYRPTKRVKPMSEIMI  
DAGRIITDVIDKELIOPEFIEIVGVEYIARAGWPEKKAWSILLDSORLOIWLIS  
PELPIPEDDVMERQRLIESVPDSVTPLLIYEETDIIWNIHDIHFVFPQSHSEIER  
IFASECTGFRHLKITSI"  
BASE COUNT 856 a 562 c 615 g 745 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,41e-233 Length: 2778  
Score: 2455.00 Matches: 464  
Percent Similarity: 99.78% Conservative: 0  
Best Local Similarity: 99.78% Mismatches: 1  
Query Match: 98.91% Indels: 1  
DB: 9 Gaps: 0  
US-10-070-464-5 (1-465) x AK000290 (1-2778)  
Qy 1 ThrGlyThraAlaAsnProLySValThrPheLySweSerGluIleMetlleAspAlaGlu 20  
Db 1114 ACAGTACACCAATCTTAAGTCACTTTTAAAGTGCAGAAATTAATGATGCTGAA 1173  
Qy 21 GlyArgllelleAspVallleAspLySglUleuIleGlnProPheGluIleLeuPheGlu 40  
Db 1174 GGAAGCATCATAGATGTCATATAGATTAAGAACTAATTCACCTTTGAGATTTCTTGA 1233  
Qy 41 GlyValGluLyrrlleAlaArgAlaGlyTrrThrProGluGlyLySValATPrrSerlle 60  
Db 1234 GGAAGTGAATATATATGCAAGCTGATGATGATCTCTGAGGAAATATGCTGCTCATC 1293  
Qy 61 LeuLeuAspArgSerGlnTrArgLeuGlnlleValLeuIleSerProGluLeuPheIle 80  
Db 1294 CTACTAGATGCTGCCAGACTCCGCTCAATAGTGTGATCTCACTGAATTTATATC 1353  
Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
Db 1354 CCAGTAGAAGATGATGTTATGAAAGCAGACAGACTGATGATGATGATGATGATGATG 1413  
Qy 101 ThrProLeuIlelleYrGluGluGluThrThaAspIleTrrPheAsnIleHisAspIlePhe 120  
Db 1414 ACCCACTAATTTATATGAAGAAACACAGACATCTGATTAATATCATGACATCTTT 1473  
Qy 121 HisValPheProGlnSerHisglUglUglUlleGluPheIlePheAlaSerGluCysVal 140  
Db 1484 CATGTTTTTCCCAAGTCCAGAAAGCAATTAAGTTTATTTTGGCTCTGAATGCAAA 1533  
Qy 141 ThrGlyPheArgHisleuYrlySleThrSerlleLeuLySglUserLySlyrLyArg 160  
Db 1534 ACAAGTTTCGTCATTTATTAACAATAATTAATCAT-TTAAAGGAAGCAATTAACGA 1592  
Qy 161 SerSerGlyLyLeuProAlaProSerAspPheLySProIleLySglUglUlleAla 180  
Db 1593 TCCAGTGGTGGCTGCTGCTCCAGAGATTTCAAGTGTCTTCAAAAGAGGATGACA 1652  
Qy 181 IleThrSerGlyGluTrpGluValLeuGlyArgHisglYserAsnIleGlnValAspGlu 200  
Db 1653 ATTACCAAGTGTGAATGGAAAGTTCTTGGCGGCAAGATGATTAATTAACAAGTGA 1712  
Qy 201 ValArgArgLeuValTyrrPheGluGlyThrLyAspSerProLeuGlnHisSleuTyrr 220  
Db 1713 GTCAGAAAGCTGTATATTTTGAAGGACCAAAAGACTCCCTTTAAGCATCACTGTAC 1772  
Qy 221 ValValSerTyrrAlaAsnProGlyGluValThrArgLeuThrAspArgGlyTyrrSerHis 240  
Db 1773 GTAGTCAGTTACCTTAATCTGAGAGAGTGAACAAGCTGACTGACCGTGGCTACTCAT 1832  
Qy 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLySlyrSerAsnGlnLyAsn 260

Db 1833 TCTTGGCGCATCAGACAGACTGAGCTTTTATTAAGTATAGTATACAGAGAT 1892  
 Qy ProHisCysValSerLeuTyrIleuSerSerProGluAspAspProThrCysIleThr 280  
 Db 1893 CCACACTGTGTCTCCCTTACAGATCAAGTCTGGAAGATGACCCAACTGGCAAAACA 1952  
 Qy 281 LysGluPheTyrPalaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 1953 AAGGAATTTGGGACACCATTTTGGATTGAGAGAGTCTCTTCCGACTATACCTCTCCA 2012  
 Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIlePro 320  
 Db 2013 GAAATTTCTCTTGGAAAGTACTAGATTTTACATTTGATGGGATCTCTACAGGCT 2072  
 Qy 321 HisAspLeuGlnProGlyIleLeuSerTyrProThrValLeuPheIleTyrGlyProGln 340  
 Db 2073 CATGATCTACAGCCCGGAAAGAAATATCTACTGCTGCTCATATATGCTGCTCTAG 2132  
 Qy 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 Db 2133 GTTGCTATTGCTGGGGCCCGACATCTGCTGATCTTCTATGATACAGATACAGGAA 2192  
 Qy 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 Db 2193 CGTATATAGGTCACCTGACCAAGATACAGGAGCTATTAAGGATCTGTGGCCATG 2252  
 Qy 381 GlnAlaGlyIlePheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 Db 2253 CAACGAGAAAGTTCCCTCTGACCAATCGTTTACTGCTTATCATGCTTCTCTGAT 2312  
 Qy 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIle 420  
 Db 2313 GAGATATGCCATTTTGGACATACCGAGTATATTACGAGTATTTTAAAGTACGCGCTGAAAG 2372  
 Qy 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 Db 2373 CCAATGATTTTACAGATTTATCTCTCAGAGAGACACACATTAAGGTTCTCGAANTCGGGA 2432  
 Qy 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 Db 2433 GAACATTATGAACGATCTTTTGGACTACCTTCMAAGAAACCTTGATCATCGTATTCCT 2492  
 Qy 461 AlaLeuLysValIle 465  
 Db 2493 GCTCTAAAGTGATA 2507  
 RESULT 5 3130 bp mRNA linear PRI 30-MAY-2002  
 LOCUS BC030688  
 DEFINITION Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:26191  
 IMAGE:4822550, mRNA, complete cds.  
 ACCESSION BC030688.1 GI:21265132  
 VERSION MGC.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens; Chordata; Craniota; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 3130)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>  
 contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha  
 Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>  
 Series: IRAC Plate: 33 Row: d Column: 5  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 18450277.

## FEATURES

## source

1..3130  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:26191 IMAGE:4822550"  
 /tissue\_type="Testis"  
 /clone\_1db="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="vector: pBluescript"  
 372..2867  
 /codon\_start=1  
 /product="similar to dipeptidylpeptidase 8"  
 /protein\_id="AAH30688.1"  
 /db\_xref="GI:21265133"  
 /translation="MAAAMETEQQLVEIFETADCEENIESQDRPKLEPFYVERYSMSQ  
 LKRLADTRKHYGDMKAPRDPFVRVNDGSHSRITYLAMSGEREXTLPYSGE  
 PKTNRNAVLMISKRELLDFOALDYGNDGSRRELLRERRIGTVGASVDYHSG  
 TFLFOASGCIYHVDGPGPGFTQPLRPNLVETSCPINRDPKCRPADPMAIFATHSN  
 DIWISNITVREERLITVHNELAMNEEDASAGATVLOEEFVRSYSCWPCPAEET  
 PSGGKILRIILENDESEVEIIVHTSPLMERADSRFRYPTGTANPVTKMEIMT  
 DAEGRIDIVIDKILQPEILFEGVEIYIAGMPEKAYMSILDSQTLQVLIS  
 PELFIPEDVDMERORLIESVDSVTPLIYEETDIDMINHDIFHFPQSHSEIER  
 IPASECKTGRHLKYKINSILKESKYRSGGLPAPSPKCPKIEAIIITSGEMVLGR  
 HGSNIQVDERRLVYFEGTQSPLEHLLVYVSYNRPSEVRLTRGRSHSCICQHC  
 FFIKSYSNQKPHCVSLYKLSPEDDPCTKREKRWATLLDSAGLPDTPETSPES  
 TTGTLTGLGMXKRPDLOPGKRYPTVLFIYGGPOVLNNRKGVKRYRMLNLAIGVY  
 VVVIDNRGSKHRGLKPEGAFFKYKVAIAGAAPTLMIFYDGTGYERVMGHDPQNEGYV  
 LGSVAMQAEKFPSEPNLLHGLFDENVHFATSIILSFVLRAGKPYDQIYPQERH  
 SIRVSEGEHYEHLHLHYLOENIGSRITALKVI"  
 BASE COUNT 955 a 626 c 717 g 832 t  
 ORIGIN

## Alignment Scores:

Pred. NO.: 9,27e-233 Length: 3130  
 Score: 2447.50 Matches: 465  
 Percent Similarity: 90.47% Conservative: 0  
 Best Local Similarity: 90.47% Mismatches: 49  
 Query Match: 98.61% Indels: 49  
 DB: Gaps: 1

US-10-070-464-5 (1-465) x BC030688 (1-3130)

Qy 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 1323 ACAGGTACAGCAATCTTAAGTCACTTTAAAGTGCAGAAATAAGATGAGCTGAA 1382  
 Qy 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 1383 GGAGAGTCATGATGTCATGATGATGAGAACTAATTCACCTTTTGAGATTTCTATGAA 1442  
 Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 Db 1443 GGAAGTTAATATATATGCGCAAGCTCGATGCACTCCTGAGGGAATAATGCTGTCATC 1502  
 Qy 61 LeuLeuAspLysSerGlnTyrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1503 CTACTAGATGCTCCACAGCTCCCTTACAGATGATGTGATCTTCACTGAATTAATTATC 1562  
 Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1563 CCACTGAGAATATATGTTATGAAAGGCAAGACTCATTTAGTCAGTGGCTGATTCGTG 1622

OY	101	ThrProLeuIleIleIeTyGluGluThrThrAspIleIleTrpIleAsnIleHisAspIlePhe	120
Db	1623	ACGCCACTAATATTATCTATGAAGAAACAAGACACTGTGAAATTAATCTCAAGACATCTTT	1682
OY	121	HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIleAs	140
Db	1683	CATGTTTTCCTCCAAAGTCACGAAAGGAATATGATTTATTTTGGCTCTGAATGCAA	1742
OY	141	ThrGlyPheArgHisLeuTyIleAsnIleThrSerIleLeuIleGluSerIleGlyIleAsn	160
Db	1743	ACAGGTTTCCTCGCATTTATACAAATTTCACTATTTTAAAGAAAGCAAAATATAAACGA	1802
OY	161	SerSerGluGlyLeuProAlaProSerAspPheIleCysProIleGluGluGluIleAla	180
Db	1803	TCCAGTGGGGCTGCTCCTGCTCCAAAGTAATTCAAAGTCTCTATCAAGAGAGATACGA	1862
OY	181	IleThrSerGluGluTrpGluValLeuGluIleArgHisGluSerAsnIleGluValAspGlu	200
Db	1863	ATTACACAGTGGGAATGGGAAGTTCTTGCGCGGCATGATCTAATATCCAAAGTTGATGAA	1922
OY	201	ValAlaGluGluLeuValTyrPheGluGlyThrIleAspSerProLeuGluHisHisLeuTyr	220
Db	1923	GTCAGAAAGCTGGTATATTATTGAAAGCAACAAAGCTCCCTTTAGACATCACCTGTAC	1982
OY	221	ValValSerTyrValAsnProGluGluValThrArgLeuThrAspArgGlyTyrSerHis	240
Db	1983	GTAGTCAGTTACGTAAATCTCGAGAAAGGTGACAAAGCTGACGTGACCGTGGCTATCACT	2042
OY	241	SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGluIleAsn	260
Db	2043	TCTTGCTCATCAGTCACACACTGTGACTCTTTTAAAGTAAGTATAGTAAACCAAGAT	2102
OY	261	ProHisCysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIleThr	280
Db	2103	CCACACTGTGTCTCCCTTTACAGGTATCAAGTCTGAAAGATGACCAACTGGCAAAACA	2162
OY	281	LysGluPheTrpAlaThrIleLeuAspSerAlaGluProLeuProAspTyrThrProPro	300
Db	2163	AAGGATTTTGGGCCACCATTTTGGATTACAGAGGCTCTTCTCGAATATATCTCTCA	2222
OY	301	GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyIleLeuTyrIlePro	320
Db	2223	GAAATTTTCTCTTTTGAAGATACACTGATTTAATTGTATGGAGTCTCTACAAAGCT	2282
OY	321	HisAspLeuGlnProGluIleIleIleTyrProThrValLeuPheIleTyrGlyIleProGln	340
Db	2283	CATGATCTACAGCCTCGGAAGAAATCTCTACTGTGCTTCAATATAGTGGTCTCTCAG	2342
OY	340	-----	340
Db	2343	GTGCAGTTGTGAATATACGTTTAAAGAGTCAAGATTTCCGCTTGAAATACCCTAGCC	2402
OY	340	-----	340
Db	2403	TCTCTAGGTTATGTGTTGTGTAGTATACAAACAGGGATCTCTGACCGAGGCTTAAA	2462
OY	341	-----ValAlaIleAlaGlyAlaProValThrLeuTrp	351
Db	2463	TTTGAGAGCGCTTTAAATATATAAATAGTGTCTATTGCTGGGGGCCACAGTCACTGTGG	2522
OY	352	IlePheTyrAspThrGlyTyrThrGluArgTyrIleGlyHisAspAsnGluIleAsnGluGln	371
Db	2533	ATCTTCTATGATACGATACCGGATACGATATATATGGGTCTACCTGTACGAGAAATGAACG	2582
OY	372	GlyTyrTyrLeuGlySerValAlaMetGlnAlaGluIlePheProSerGluProAsnArg	391
Db	2583	GGCTATTACTTAGATCTGTGGCCATGCAACAGAAAGTCCCTCGTGAACCAATCGT	2642
OY	392	LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisIleThrSerIleLeu	411
Db	2643	TTACTGCTTACACATGTTTCTCTGATGATGAGATGTCCATTTTGGACATACACAGATATTTA	2702

QY	412	LeuserPheLeuValaArgAlaGlyLysProCtyrAspLeuGlnIleTyrProGlnGluArg	431
DB	2783	CTGAAGTTTTTATGAGGGCTGGAAACCATATATTAACGATCTAATCTCAGAGAGA	2762
QY	432	HisSerIleAaGValProGluSerGlyGlnIleTyrGlnLeuHisLeuLeuHisTyrLeu	451
DB	2763	CACAGCATTAAGATTTCTCGAATCCGGAGAACATTAATGAATCGATCTTTTGACATCCTT	2822
QY	452	GlnGlnLeuLeuGlySerArgIleAlaAlaLeuLysValIle	465
DB	2823	CAAGAAACCTTGGATCACTATTGCTGCTCTAAAGAGATA	2864
RESULT 6			
LOCUS	AX608731	4523 bp	DNA
DEFINITION	Sequence 8 from Patent WO0231134.		linear
ACCESSION	AX608731		
VERSION	AX608731.1	GI:28404301	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE			
AUTHORS	Qi, S., Akineanya, K.O., Riviere, P.J. and Junien, J.L.		
TITLE	Novel serine protease genes related to dppiv		
JOURNAL	Patent: WO 0231134-A 8 18-Apr-2002;		
FEATURES			
source	location/Qualifiers		
	1..4523		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	1384 a 828 c 940 g 1371 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,93e-232	Length:	4523
Score:	2446.50	Matches:	463
Percent Similarity:	99.57%	Conservative:	0
Best Local Similarity:	99.57%	Mismatches:	1
Query Match:	98.57%	Indels:	2
DB:	6	Gaps:	1
US-10-070-464-5 (1-465) x AX608731 (1-4523)			
QY	1	TherGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu	20
DB	1165	ACAGGTACAGCAATCTCTAAAGTCACTTTTAAGTGCAGAAATAAGATTGAGTGCAGA	1224
QY	21	GlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGlnIleLeuPheGlu	40
DB	1225	GGAAAGATCATAGATGTCATATGATPAAGAACTAAATTCACCTTTTGAGATTCTATTGAA	1284
QY	41	GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle	60
DB	1285	GGAATTGAATATATATGCGACAGCGTGGATGGAATCTCTAGGGAAATATGCTGCTCCATC	1344
QY	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle	80
DB	1345	CTACTTAATGCTCTCCAGACTCGCTCAAGATAGTTGATCTCACTGAATTATTTATC	1404
QY	81	ProValGluAspAspValMetGluArgArgGlnIleGluSerValProAspSerVal	100
DB	1405	CCATTAGAAAGTATGTTATGGAAGGACAGATCTATTAGTCAAGCTGATTCGTG	1464
QY	101	ThrProLeuIleIleTyrGlnGluThrThrAspIleTyrPheAsnIleHisAspIlePhe	120
DB	1465	ACGCCACTAATTATCTATGAAGAAACAACAGACATCTGGATTAATATCCATGACATCTTT	1524
QY	121	HisValPheProGlnSerHisGlnGluGlnIleGluPheIlePheAlaSerGluCysLys	140
DB	1525	CATGTTTTTCCCAATCAACGAAGAGAAATGAGTTATTTTTTGCTCTGAATGCAGA	1584

QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 DB 1585 ACAGGTTTCGCTCATTTATACAAATATCATCTATTTAAAGGAACAAATATAACGA 1644  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluLysIleAla 180  
 DB 1645 TCACAGTGGTGGCTGCTGCCAAGATTTCCAAAGTGCCTTATCAAAAGGAGATGCA 1704  
 QY 181 IleThrSerGlyGluTTPGluValIleGluIYArgHisGlySerAsnIleGluValAspGlu 200  
 DB 1705 ATTACCGATGCTGATATGGAAGTTCTTGCCCGCATGATCTAATATCCAAAGTTGATGA 1764  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220  
 DB 1765 GTCAGAAAGCTGGATATTTTGAAGGACCAAAAGACTCCCTTAAAGCATCACTGTAC 1824  
 QY 221 ValValSerTyrValAspProGlyGluValIleThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 1825 GTAGTCAGTACGTAATCCTGAGAGGTACAAAGCTGACGACCTGCTACTCAT 1884  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 DB 1885 TCTTCTGCTGATCGATCGACCTGACTCTTTTATAGTAAGTAAGTAACCAAGAAAT 1944  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 1945 CCACACTGTGTGCTCCCTTACAAAGCTATCAAGTCTGAAAGTGAACCACTGCAAAACA 2004  
 QY 281 LysGluPheThrAlaThrIleLeuAspSerIleGlyProLeuProAlaTyrThrProPro 300  
 DB 2005 AAGGAATTTTGGGCGACCATTTTGGATTGACAGAGTCTCTTCTGACTATCTCCCA 2064  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLysLysPro 320  
 DB 2065 GAATTTTCTCTTTGAAAGTACTACTGATTTACATTTGATGGATGCTCTACAGACT 2124  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGln 340  
 DB 2125 CATGATCTACAGCTCGAAAGAAATATCTACTGCTGCTGCTATATATGCTGCTGCG 2180  
 QY 341 ValAlaIleAlaGlyAlaProValIleLeuThrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 2181 GTTGCTATTCCTGGGCGCCAGTCACTCTGTGATCTTCTATGATACAGATACACGGA 2240  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 2241 CGTTATATGGGTACCCCTGACCAAGATGAACAGGCGTATTTAGGATCTGTGGCATG 2300  
 QY 381 GlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 DB 2301 CAAGCAGAAAGTTCCCTCTGAAACCAATCGTTTACTGCTTACATGCTTCCCTGAT 2360  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLys 420  
 DB 2361 GAAATATGCTATTTGACATACCAATATATCTAGTGTATTTTGAAGAGGCTGGAAG 2420  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 DB 2421 CCATATGATTTTACAGATCTATCTCCAGAGAGACACAGCATTAAGATTTCTGATCGGA 2480  
 QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 DB 2481 GAACATTAATGAACTGACTCTTTTGGACTACTCTTCAAGAAACCTTGATCATGATTCCT 2540  
 QY 461 AlaLeuLysValIle 465  
 DB 2541 GCTCTAAAGTGATA 2555  
 RESULT 7  
 BD157001 2161 bp DNA linear PAT 17-JAN-2003  
 LOCUS BD157001  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.

ACCESSION BD157001  
 VERSION BD157001.1 GI:27862759  
 KEYWORDS JP 2002191363-A/11844-  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 2161)  
 AUTHORS Oca,T., Isoqai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
 TITLE Primer for synthesizing full-length cDNA and use thereof  
 JOURNAL Patent: JP 2002191363-A 11844 09-JUL-2002;  
 COMMENT HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/11844  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOQAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC 10,  
 C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 FEATURES  
 Source Location/Qualifiers  
 1..2161  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 FT CDS location(116)..(1911).  
 BASE COUNT 662 a 429 c 452 g 618 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2..2e-232 Length: 2161  
 Score: 2441.50 Matches: 464  
 Percent Similarity: 90.27% Conservative: 0  
 Best Local Similarity: 90.27% Mismatches: 1  
 Query Match: 98.37% Indels: 49  
 DB: 6 Gaps: 1  
 US-10-070-464-5 (1-465) x BD157001 (1-2161)  
 QY 1 ThrGlyThrAlaAspProLysValIlePheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 370 ACAGGTACAGCAATCTTAAAGTCACTTTAAGATGTCAAGAAATATGATTTGATGCTGAA 429  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 430 GGAAGATCATGATGATTCATAGTAAAGAACTAATTCACCTTTTGAGATTCATATTGAA 469  
 QY 44 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 DB 490 GGAGTTGAATATATTTCCAGAGCTGATGAGACTCTGAGGAAATATATGCTGTGCATC 549  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 550 CTACTAGATGCTCCCGACCTGCTTACAGATAGTGTATTCACCTGATTAATTTATTC 609  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 610 CCAGTGAAGATGATGTTTATGGAAGGACAGACTCATTAAGTCACTGCTGATTTCTGTG 669  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 670 ACCGCACATAATATCATATGAGAAACACAGACATCTGATTAATATTCATGACATCTTT 729  
 QY 121 HisValPheProGlnSerHisGlyGluGluGluIleGluPheIlePheAlaSerGlyCysLys 140  
 DB 730 CATGTTTTCCTCCAAAGTACAGAGGAAATGAGTTATTTTTCCTCTGATTCGAA 789

Qy 141 ThrGlyPheArgHisLeuTyrLeuTyrLeuSerIleLeuLysGluSerLysTyrLysArg 160  
 Db 790 ACAGGTTTCCTCATTTTATACAAAATTCATCTATTTTAAAGAAAGCAATATTAAGCA 849  
 Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 Db 850 TCCAGTGGTGGGCTGCCTGCCAAGGATTCATTCATCAAAAGAGAGATACCA 909  
 Qy 181 IleHisSerGlyGluTTPGluValLeuGlyArgHisGlySerAsnIleGluValAspGlu 200  
 Db 910 ATTACCGATGGTGAATGGGAAGTCTTGCCGCGATGATATATATCCAAAGTTGATGAA 969  
 Qy 201 ValArgArgLeuValIlyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220  
 Db 970 GTCAAGAGCTGATATATTTGAAAGCCAAAGATCTCCCTTTAGACATCACCTGTAC 1029  
 Qy 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1030 GTAGTCAGTACGTAATCTCGAGAGGTGACAAAGCTGACCTGACCGTACTACAT 1089  
 Qy 241 SerCysCysIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGluLysAsn 260  
 Db 1090 TCTTGCTGCATCAGCTCAGCAGCTGTGACTCTTTATAGTATAGTATACAGAGAT 1149  
 Qy 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 Db 1150 CCACACGTGTGTCTCTTTACAGCTATCAAGTCTGAAAGTGAACCCAACTTGCAAAACA 1209  
 Qy 281 LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 1210 AAGGAATTTGGGCGACCATTTTGATTCACAGAGTCTCTTCCGACTATATCTCTCA 1269  
 Qy 301 GluIlePheSerPheGlySerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 Db 1270 GAAATTTCTCTTTGAAAGTACACGATTTTACATTGATGAGAGCTCTTACAGGCT 1339  
 Qy 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 Db 1330 CATGATATACAGCTGGAAGAAATATCTCTGTGCTCATATATGAGTGTCTCAG 1389  
 Qy 340 ----- 340  
 Db 1390 GTGCGATTGGTAATAATCGGTTTAAAGAGTCAAGTATTTCCGCTGAATACCTTAA 1449  
 Qy 340 ----- 340  
 Db 1450 TCTCTAGGTTATGTGTGTAGTAGACACAAGGAGATCTGTCAACGAGGCTTAAA 1509  
 Qy 341 -----ValAlaIleAlaGlyAlaProValThrLeuTyr 351  
 Db 1510 TTTGAAGCGCCCTTAATAATATAAAATGTTCTATTGCTGGGCCCCAGTCACTCTGTGG 1569  
 Qy 352 IlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGln 371  
 Db 1570 ATCTTCATATACAGATACACGAAAGCTATATGAGTCACTGACCAAGATGAAGACAG 1629  
 Qy 372 GlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAspArg 391  
 Db 1630 GGCATATCTTAGATCTGGGCGATGACAGCAAAAGTCCCTCGAAGCAAAATCGT 1669  
 Qy 392 LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 411  
 Db 1690 TTAAGTCTCTTACATGTTCTCTGATGAGAGATTCATTTTGACATACAGATATTTA 1749  
 Qy 412 LeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArg 431  
 Db 1750 CTAGATTTTTAGAGAGGCTGGAAGCCATATGATTTACAGATCTATCTCCAGAGAGAGA 1809  
 Qy 432 HisSerIleArgValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeu 451  
 Db 1810 CACAGCATTAAGAGTTCTGTAATCGGAGAACATTTATGAATCTGCAATCTTTTGCACTACCTT 1869

Qy 452 GlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 465  
 Db 1870 CAGAGAAACCTTGATCACGATATTCCTCTTAAAGAGTAA 1911

RESULT 8  
 AK027826  
 LOCUS  
 DEFINITION Homo sapiens cDNA FLJ14920 fis, clone PLACE1007416, weakly similar to D1EPTDYL PEPTIDASE IV (EC 3.4.14.5).  
 AK027826 2161 bp mRNA linear PRI 01-AUG-2002  
 AK027826.1 GI:14042789  
 ACCESSION  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS  
 1 Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Magatsuma,M., Hosokari,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y., Nimomiya,K. and Iwayanagi,T.  
 NEBO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 2161)  
 Isegai,T. and Otsuki,T.  
 Direct Submission

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 1 Submitted (10-MAY-2001) Takao Isegai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986)  
 NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

COMMENT  
 FEATURES  
 source  
 1. 2161  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="PLACE1007416"  
 /cissue\_type="placenta"  
 /clone\_id="PLACE1"  
 /note="cloning vector: pME18SFLJ3"  
 16..1914  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB55395.1"  
 /db\_xref="GI:14042790"  
 /translation="MDPKLCPADPDMIAFIHSNDIWNISIVTREERLTYVHNELANN EEDARSAGVATFVLOEERDYSGYMCPKABETPSGKILRIYEENDESVEIIVHT SPMLERBASFPYKGTANTPVTPEKMSIMIDABERIIIDVDEKELIOPETILPECV EYLARAKMTPEBKYAMSLIDRSQTRQIVLISPELPIPEDDVMQRQLTESPDSY TPLILBETDIDMINIHIDIFHVPQSEIEIFLIPASECTGFHLKITSILKESY KRSSGGLPAPSDKCPKIEIATISGEWVLGRGSMIOVDEVRLVFPSTGKSPLE HHCYVVSYPADSEVTRITDRSYSHSCISQCHDFISIKYOKRPHCVSYKXSSPED DPCKTKEFWATILDSAGLPDYTPPIPFSESTGFTLYGLMKLPDLOPGKKYPV LPIYGGPOVOLVNNRPKGVYKPLNTLASIGYVVVVDNNGSCHRGKPGKAPYKPV AINAGPVTLMIFYDTGTERYMHGPDNDEGYILGSAVMAEKRPSPNLLHLHGL DENVHRAHTLSILSPFVRACKPYDLOLYPQERSIRVPSGEHYEHLHLHYLDENLDS RIALAKVI"

## CDS

BASE COUNT 662 a 429 c 452 g 618 t  
 ORIGIN  
 Alignment Scores: \*  
 Pred. No.: 2,2e-232 Length: 2161  
 Score: 2441.50 Matches: 464  
 Percent Similarity: 90.27% Conservative: 0  
 Best Local Similarity: 90.27% Mismatches: 1









/product="dipeptidyl peptidase IV-related protein-1"  
/protein\_id="AA017261.1"  
/db\_xref="GI:27549550"  
/translation="MAAMETEOGLVEIFETADCEENIESODRPLEPFYVERYSMSO  
LKLIADTRKYHGMAKAPHDPMFVRKNDPGHSPRIYVILAMSGENRENTLPSYRI  
PKTINRAVIMLSMKPLLDLPQATLDYGMYSREBELREKRGITVIAIADYHQSOG  
TFLPQAGSGIYHVADGGPQPTQPLPNLIVDSCPNIDPVCMDPMLAIFSHN  
DIWISNVTREERLTIVNHELAMEEDASAGVATVLOEEDRSGYWCPEAETP  
PSGKILRLIYEENDESEVEIIVTSPMLSTRADSRFRYPKGTANPKYFKMSIWI  
DAEGRIIDVIDKEILOPELIFEGVEYIAGMTPEGVKAMSILLDSOTRLVLS  
PELPIVEDDMERORLIESVPSVPLIYEETDOWIMIHIFHVPQSHSEELER  
IFASECTGRFRLIKTSIIKESYKRSSSGLPAPSPKPIREIATISGENBVLOR  
HGSNIQDEVRRLVYFEGIKOSPLEHHLIVVSVNRECVRLIDRGISHSCISQHD  
FPISKYSNQNPHCVSLYKLSPEDDPTCKEFWATIILOSAGLPDYTPPEIISFES  
TFGTLYGMLYKPHDLOPKKYPVLVLIYGQVQLVYNNRFGKYPRLNTLASLGY  
VVVIDNRGSCRGKFEKAFYKMGQIIEIDOVEGLQYLSRYDFIDLDHGMGYS  
GGLSLMAIMORSDFRVALIAGAVPTLMFYDGYTERVGHDPQNGYGLVSAMQ  
AKRPEPNRLIIHGFLENVPHAKTSIILSFLVRAGKDYDQIYRQBNHSIRVES  
GEHYELHLHYLDENLGSRIALKVT"  
BASE COUNT 804 a 517 c 588 g 747 t  
ORIGIN

## Alignment Scores:

Pred. No.:	2,52e-230	Length:	2656
Score:	2422.00	Matches:	465
Percent Similarity:	82.30%	Conservative:	0
Best Local Similarity:	82.30%	Mismatches:	0
Query Match:	97.58%	Indels:	100
DB:	9	Gaps:	1

US-10-070-464-5 (1-465) x AY172659 (1-2656)

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
DB 959 ACAGGTACACCAATCTTAAGTCACCTTTAAGATGCAGAAATTAAGCTTGAATGCAGAA 1018  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
DB 1019 GGAAGGATTCATAGATGCATAGATTAAGAACTAATTCACCTTTGAGATTCATATTGAA 1078  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTSPSerIle 60  
DB 1079 GGAAGTGAATATATATGCGCAAGCTGATGACCTCTGAGGAAATATGCTTGCTCATC 1138  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1139 CTACTAGATCGCTCCAGACTCGCTGAGATGATGTTGATCTCACTGAATATTATTATC 1198  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1199 CCAGTAGAAAGATGATGTAAGAAAGCAGAGACTGATGATGATGAGCTGATTCGTG 1258  
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
DB 1259 ACCCCACTATTTATCTATGAAGAAACAAGACATCTCGATTAATATCCATGACATCTTT 1318  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1319 CATGTTTTTCCCAAGTCACAGAAAGAGAAATGAGATTATTTTGGCTCTGAAATGCAAA 1378  
QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
DB 1379 ACAGGTTTCCGTCATTTATACAAATATACATCTATTTTAAAGAAAGCAATATTAACGA 1438  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluIleAla 180  
DB 1439 TCAGTGGTGGGCTGCTGCCAAGATTTTCAAGTGTCTCTTCAAAAGGAGATGCA 1498  
QY 181 IleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1499 ATTACAGTGTGTAATGGGAAGTCTTGCCGCGCATGATCTAATATCCAAAGTTGATGAA 1558  
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluIleHisLeuTyr 220

DB 1559 GTGAGAAAGCTGGATATTTTGAAGGACCAACCAAGACTCCCTTTAGAGCATCACCTGTAC 1618  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
DB 1619 GTAGTCATGATGATTAATCTCGAGAGGTACACAGGCTGACCTGCGGTACTTCAAT 1678  
QY 241 SerCysLeuIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGlnLysAsn 260  
DB 1679 TCTTGCGCATCAGTACGACCTGACCTCTTTATAGTAAAGTATATCTAACCAGAAAT 1728  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1739 CCACACTGTGTGCTCCCTTCAAGACTATCAAGCTCGAAGATGACCCCACTTCGCAAAA 1798  
QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
DB 1799 AAGGAATTTTGGGCGCACATTTTGGATTCAGAGGCTCTTCTGACTATACCTCTCCA 1858  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 1859 GAATTTTCTCTTTGAAAGTACTACTGATTAATGATGAGATGCTTCAACACCT 1918  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
DB 1919 CATGATCTACAGCTCGAAGAAAGAAATCCATCTGTGCTTCAATATGTGTGCTCAG 1978  
QY 340 ----- 340  
DB 1979 GTGCACTGTGATTAATGATTAAGAGCAAGTATTTCCGCTTGAATACCCTAGCC 2038  
QY 340 ----- 340  
DB 2039 TCTCTAGTATGTTGGTTGATGATAGACACAGGGGATCTGTACAGAGGCTTAA 2098  
QY 340 ----- 340  
DB 2099 TTTGAAGCGCTTTAAATATTAATGGGTCAATAGAAATGACGATCAGTGGAGCA 2158  
QY 340 ----- 340  
DB 2159 CTCGATATCTACTCTTGATATGATTTGATTCATGATGCTGTGGCATCCAGGC 2218  
QY 340 ----- 340  
DB 2219 TGTGCTTAGAGGATACCTCTCCGTGATGATTAATGACAGAGCTCAGATCTTCAGG 2278  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
DB 2279 GTTGCTATGCTGGGCGCCAGACCTGCTGTGATCTTCTATGATACAGATACACCGAA 2338  
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
DB 2339 CGTTATATGGGTACACCTGACCAAGATGAACAGAGGCTATTAATCTTGAGATCTGGGCCATG 2398  
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
DB 2399 CAAGCAGAAAAGTTCCTCTGAAACCAATCGTTTCTGCTTCAATGCTTTCCTGAT 2458  
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
DB 2459 GAGAAATGTCATTTTGCATACATACAGATATATTAAGTATTTTGAAGAGGCTGGAAG 2518  
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
DB 2519 CCAATGATTTACAGATCTATCTCTCAGAGAGACACAGCTTAAGATTTCTGATCGGA 2578  
QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
DB 2579 GAACATTAAGAACTGATCTTTTGGACCTACCTTCAAGAAAACCTTGATCAGATATGCT 2638  
QY 463 AlaLeuLysValIle 465



Db 2579 GAACATTATGACATGCACTTTTGGACTCTTCAAGAAACCTTGATCAGATTGCT 2638  
Qy 461 A|A|e|u|y|s|v|a|l|l|e| 465  
Db 2639 GCTCTAAAAGTGATA 2653  
RESULT 12  
AX342633  
LOCUS AX342633 3106 bp DNA linear PAT 12-JAN-2002  
DEFINITION Sequence 30 from Patent WO0198468.  
ACCESSION AX342633  
VERSION AX342633.1 GI:18152030  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Yue H., Elliott V.S., Gandhi A.R., Lal P., Au-Young J.,  
Tribouley C.M., Deleage A.M., Baughn M.K., Nguyen D.B., Lee E.A.,  
Hatala A., Khan F.A., Walla N.K., Iao M.G., Lu D.A., Patterson C.,  
Tang Y.T., Walsh R.T., Azimzal Y., Ramkumar J., Xu Y. and Reddy R.  
JOURNAL Patent: WO 0198468-A 30 27-DEC-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
source 1..3106  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7160544CB1"  
BASE COUNT 928 a 633 c 704 g 841 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,13e-230 Length: 3106  
Score: 2422.00 Matches: 465  
Percent Similarity: 82.30% Conservative: 0  
Best Local Similarity: 82.30% Mismatches: 0  
Query Match: 97.58% Indels: 100  
DB: Gaps: 1  
US-10-070-464-5 (1-465) x AX342633 (1-3106)  
Qy 1 ThrGlyThrAlaAspProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
Db 1154 ACAGGTACAGCAAAATCTTAAGTCACTTTAAGATGTCAGAAATAAGATTGATGCTGAA 1213  
Qy 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
Db 1214 GGAAGGATCATAGATGTCATGATTAAGAACTAATTCACCTTTGAGATTCTATTGAA 1273  
Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTTPSerIle 60  
Db 1274 GGAATTAAATATATGTCAGAGCTGATGAGCTCCTGAGGAAATATGCTTGCTCATC 1333  
Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1334 CTACTAATAGCTCTCCAGACCTCGCTACAGATAGTGTGATCTCACTGAATTTATTATC 1393  
Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
Db 1394 CCAAGTAGAATGATGTTATGAGAAAGCAGAGACTCATTTGAGTCAGTCCGATTCGTG 1453  
Qy 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrPheLeuIleHisAspIlePhe 120  
Db 1454 ACCGCACTAATATCTATGAAAGAAACAACAGACATCGAATAATATCCATGACATCTTT 1513  
Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
Db 1514 CATGTTTTTCCCAAGTCAAGAACAGAAAGAAATTGATTTTGGCTCTGAAATCGAA 1573  
Qy 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160

Db 1574 ACAGTTTCCGTCACTTTATACAAATTAACATCTATTTTAAAGAAAGCAAAATATAACGA 1633  
Qy 161 SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
Db 1634 TCCAGTGTGGGCTGCTGCTCCCAAGTATTTCAAGTGTCTTATCAAAAGAGATAGCA 1693  
Qy 181 IleThrSerGlyGluTyrPheGluValLeuGlyValArgHisGlySerAsnIleGlnValAspGlu 200  
Db 1694 ATTACCAAGTGTGAATGGAAGATTCTGGCCGAGATGATCTAATATATCAAGTTGATGAA 1753  
Qy 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyr 220  
Db 1754 GTCAAGAGGCTGATATTTTGAAGGACCAAGACCTCCCTTGAAGACATCACTGTAC 1813  
Qy 221 ValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
Db 1814 GTACTCAGTTAAGTAATCCCGAGAGGTGACAGGCTGACTGACCGGTACTATCATCAT 1873  
Qy 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
Db 1874 TCTTGCTGCATCAGCTCAGCACTGTGACTCTTTTAAAGTAAGTATGATTAACCAAGAAAT 1933  
Qy 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspProThrCysLysThr 280  
Db 1934 CCACACTGTGTGCTCTTTACAAAGCTATCAAGTCTTAAGATGACCAACTTGCAAAACA 1993  
Qy 281 LysGluPheThrPalatThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
Db 1994 AAGGAATTTTGGGCAACATTTTGATTCAGCAGGTCCTCTTGACTATACCTCTCCA 2053  
Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
Db 2054 GAATTTTCTTTTGAAGACTACTGATGATTTACATGTATGAGATGCTCTCAACACCT 2113  
Qy 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLysProGln 340  
Db 2114 CATGATTCACGCTTGAAAGAAATATCTCATGTGCTGTTCAATATAGGTGTCAG 2173  
Qy 340 ----- 340  
Db 2174 GTGCAGTTGGTAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTAGCC 2233  
Qy 340 ----- 340  
Db 2234 TCTCTAGATTATGTGTTGATGATAGACACAGGAGATCTGTCAACGAGGCTTAAA 2293  
Qy 340 ----- 340  
Db 2294 TTTGAAGCGCCTTAAATATTAATTAAGGTCAAATTAAGAAATTGACGATCAGTGAAGCA 2353  
Qy 340 ----- 340  
Db 2354 CTCCAATATCTAGCTTCTCGATATGATTTATGACTTAGATGATGCGTGGCATCCAGGC 2413  
Qy 340 ----- 340  
Db 2414 TGGTCTATGAGAGATACCTCTCCGATGACATTAATGACAGAGTCAGATATCTTACAG 2473  
Qy 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
Db 2474 GTTGCTATTGCTGGGGCCAGATCACTCTGTGATCTTCTATATACAGATACAGGAA 2533  
Qy 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
Db 2534 CGTTATATGGGTCACTCGACCAAGATGAACAGGGCTATTAAGATCTGTGGCCATG 2593  
Qy 381 GlnAlaGluLysPheProSerGluProAspArgLeuLeuLeuLeuHisGlyPheLeuAsp 400  
Db 2594 CAAGCGAAGAAATTTCCCTGTGAACCAAAATGTTTACTGCTCTTACATGGTTCCTGAT 2653  
Qy 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
Db 2654 GAGATGTCCATTTTGCACATACCAATATTAAGATTTTGAAGGAGGCTGGAAG 2713

QY	421	ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly	440
Db	2714	CCATGTGATTACACATCTTATCTCGAGAGACACACACATGAAGTCTCGAATCGGGA	2773
QY	441	GluHisTyrGluLeuHisIleLeuHisTyrLeuGlnGlnAuaLeuGlySerArgIleAla	460
Db	2774	GACACTATGAACTGCATCTTTGGCACTACCTTGAAAGAAAACCTTGATCAGTATTGCT	2833
QY	461	AlaLeuIysValIle	465
Db	2834	GCTCTAAAGTGATA	2848
RESULT 13			
LOCUS	AF221634	3127 bp	mRNA linear PRI 05-NOV-20000
DEFINITION	Homo sapiens dipeptidyl peptidase 8 (DPP8)	mRNA, complete cds.	
ACCESSION	AF221634		
VERSION	AF221634.1	GI:11095187	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 3127) Abbott,C.A., Yu,D.M., Woollatt,E., Sutherland,G.R., McCaughan,G.W. and Gorrell,M.D.		
TITLE	Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homolog, DPP8		
JOURNAL	Eur. J. Biochem. 267 (20), 6140-6150 (2000)		
MEDLINE	20467194		
PUBMED	11012666		
REFERENCE	2 (bases 1 to 3127) Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.		
AUTHORS	Direct Submission		
TITLE	Submitted (06-JUN-2000) A.W. Morrow Gastroenterology and Liver Centre, Centenary Institute of Cell Biology and Cancer Medicine, Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia		
JOURNAL	Location/Qualifiers		
FEATURES	1..3127		
source	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="15"		
	/map="15q22"		
gene	1..3127		
	/gene="DPP8"		
	214..2862		
CDS	/note="serine protease; member of prolyl oligopeptidase 5b family belonging to enzyme clan SC; alternatively spliced forms of this mRNA are found in AF221635, AF221636 and AF221637"		
	/codon_start=1		
	/product="dipeptidyl peptidase 8"		
	/protein_id="AAC29766.1"		
	/db_xref="GI:11095186"		
	/translation="MAAAMEQLQGVLEIPEADCEENIESQDRKLEPFVVERYSWSQ LKLKLADTRKTHGYMAKAPDFMFRVKNRPNPDGSDRIYILASGNGRNTLFSBI PKTINRAVALMLSWKPLDLDFQATLDIGMYSREELRERKIDVTGIIASDYQSG TFLDQASGIVHYVKGDPGPGFTQPLRNLYVTSCPNIMRDKCPADPMWIAFHN DITWSNITVEPERLTYVHNAMERDASAGATVLOEERPRYSGYMWCRAETTT PSGKRITLILENDESEVEILHTVTSFMLERKADSRYPFGTANRKFVFKSEVIL DABSRILIDVDELLOPELLEFGEVYIARGWPEKRYAMSILDSQRLLQVILIS PELPIPEVDYMERQRLIESVDSVTPLIYEEETDITWIMHDFHYFQSHHEELIF IFASECKTGFEHLKYLITSLKESKYKSSGGLPSPDFKCPIKEIATSGEVELGR HGSNIQDEVARLYVPEGTGKSPLEHLIYVVSYPGEVTRLDRGSHSCISOHD FPLSKYSNKNPHCVSLYKLSPEPDCTKTKERWATILDSAGLPDYTPPELFSFS TGGTLYGMLYKPHDLPQKKYPTVLFTYGGPQYLVNRPFKYKTRFRLTTLASGLV VVVIDNNGCHRGLEKFGARFKMGQLEIDVQEGLOILASRDVLDLDSVGLHGSY GGYSLNALMQRSDIFPVALNAGAVTLMIFDYDTGTYERVNGHPDQNGQYGLVSVAQ AEKPPSSPNRLRLHGFIDENWVHFHNTLIFDYDTGTYERVNGHPDQNGQYGLVSVAQ GEHEHLLHYLGENTLSRIALKVI"		

BASE COUNT	943 a	637 c	706 g	841 t
ORIGIN				
Alignment Scores:				
Pred. No.:	3,16e-230	Length:	3127	
Score:	2422.00	Matches:	465	
Percent Similarity:	82.30%	Conservative:	0	
Best Local Similarity:	82.30%	Mismatches:	0	
Query Match:	97.58%	Indels:	100	
DB:	9	Gaps:	1	
US-10-070-464-5 (1-465) x AF221634 (1-3127)				
QY	1	ThrglyThrAlaAspProValThrPheLysMetSerGluLeuMetIleAspAlaGlu	20	
DB	1165	ACAGGTACAGCAATCTTAAGTCACTTTAAGATGTCAGAAATATGATGATGTA	122	
QY	21	GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu	40	
DB	1225	GGAAGACATCATGATGATGATAGTAAGAACTTAATCAACCTTTGAGATCTTATTGAA	128	
QY	41	GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle	60	
DB	1285	GGAGTGAATATATTGCCAGACGTGAAGTGAAGTCTCTGAGGGGAAATATGCTTGCTCATC	134	
QY	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleLysSerProGluLeuPheIle	80	
DB	1345	CTACTAGATCGCTCCACAGCTCGCTTACAGATAGTGTGATCTCACCTGAATATTATTC	140	
QY	81	ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal	100	
DB	1405	CCAGTAGAAGATGATGATTATGAAAGGACAGAGACTATGATGATGATGCTGATTCGTG	146	
QY	101	ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe	120	
DB	1465	ACGCACATTAATATCTATGAAAGAAACAACAACATCTGAATAAATCCATGACATCTTT	152	
QY	121	HisValPheProGlnSerHisGluGluGluTyrIleGluPheIlePheAlaSerGluCysLys	140	
DB	1525	CATGTTTTCCCAAGACACGAGAGAGAAATGATTTATTTTGCTCTGAATGCAAA	158	
QY	141	ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg	160	
DB	1585	ACAGTTTCCGCTCATTTTATACAAATTCATCTATTTAAAGAAAGCAAAATATAACGA	164	
QY	161	SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla	180	
DB	1645	TCAGTGTGTGAGCTGCTGCTCCAAAGATTTCAAGTCTCTATCCAAAGAGAGATACGA	170	
QY	181	IleThrSerGlyGluTyrGluValLeuGlyLysArgHisGlySerAsnIleGlnValAspGlu	200	
DB	1705	ATTACCAATGATGTAATGGAAGTTCTTGCCGCGATGATCTAATATCCAACTTATGAA	176	
QY	201	ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyr	220	
DB	1765	GTCGAAAGGCTGTATATTTTGAAGAGCCAAAGACCTCCCTTATAGACATCACCTGAC	182	
QY	221	ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis	240	
DB	1825	GTAGTCAGTTACGTAATCTCGAGAGGTGCAAGGCTGACGACCGGTGGCTACACAT	188	
QY	241	SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn	260	
DB	1885	TCTTGTCGATCAGTCAGCACTGACCTTCTTATAGTAACTATAGTAAACAGAGAT	194	
QY	261	ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr	280	
DB	1945	CCACACTGTGTGCTCTTTACAAAGCTATCAAGTCTGAAGATGCCCAACTTGCAAAACA	200	
QY	281	LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro	300	
DB	2005	AAGGAATTTTGGGCAACATTTTGATTCAGCAGGTCTCTTCTGACTAATCTCTCCA	206	

Qy	301	Glul1ePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro	320
Db	2065	GAATTTCTCTTCTTGAAGAGTACTAGSATTTCATTGATGAGATGCTTACAAAGCT	2124
Qy	321	HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln	340
Db	2125	CATGATCTACAGCTGGAAGAAATATCTCATCTGCTGATCATATATGATGCTCAG	2184
Qy	340	-----	340
Db	2185	GTGAGTTGGATTAATCGCTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTAGCC	2244
Qy	340	-----	340
Db	2245	TCTCTAGGTTATGTGTTGATGATAGACAAACAGGAGATCCTGTCAACGAGGCTTAA	2304
Qy	340	-----	340
Db	2305	TTTGAGGCGCTTTAAATATATAATAGGCTCAATAGAAATTGACATCAGGTGAAAGA	2364
Qy	340	-----	340
Db	2365	CTCCAAATATCTAGCTTCCGATATGATTTGATTAAGTTCGTGGGATCCAGGC	2424
Qy	340	-----	340
Db	2425	TGCTCTATGAGAGATCTCTCCCTGATGATGATTAATGACAGAGTCAATCTTCAAG	2484
Qy	341	ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGlu	360
Db	2485	GTTGCTATGCTGGGGCCCAAGTCACTGTGAGATCTTATGATACAGATACAGAA	2544
Qy	361	ArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMet	380
Db	2545	CGTTATATGGGTACCTCGACACAGATGAACAGAGGCTATTACTTAAGATCTGTGCCATG	2604
Qy	381	GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp	400
Db	2605	CAAGCAGAAAAGTCCCTCTGAAACAAATCGTTTACTGCTTACATGATGCTTCTGAT	2664
Qy	401	GluAsnValHisPheAlaHisSerSerIleLeuLeuSerPheLeuValArgAlaGlyLys	420
Db	2665	GAGATATGCAATTTTGGACATACCAAGTATATTACTGAGTTTTTAAAGAGGCTGGAAG	2724
Qy	421	ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly	440
Db	2725	CCATATGATTTTACAGATCTATCCCTCAGAGAGACACAGCATTAAGATTCGGAATCGGA	2784
Qy	441	GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnAsnLeuGlySerArgIleAla	460
Db	2785	GAACTATTATGAACGATCTTTTGGACTTACCTTCAAGAAAACCTTGATCAGTATGCT	2844
Qy	461	AlaLeuLysValIle 465	
Db	2845	GCTCTAAAAGTGATA 2859	
RESULT 14			
AX354793	3143 bp	DNA	linear PAT 06-FEB-2002
LOCUS	AX354793	Sequence 1 from Patent WO0179473.	
DEFINITION	AX354793		
ACCESSION	AX354793.1	GI:18619526	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			

source	1. .3143	/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
CDS	229..2877		
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="CAD22990.1"	
		/db_xref="GI:18619527"	
		/translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSMSQ	
		LKTLADTRKYHGMVMAKAPHDPMFVRNDPDQPSDRITYLLAMSGNRNTLVSEI	
		PKTINRAVLMLSKRPLDLFQATLDYGMATSRREBLRKRIGTGVIASTYDQSG	
		TLFLFQSGSLYHVKGDPQFTQOAPRNLYSVSCPIRNDPKLCPNDPMIATHSN	
		DIWISNIVTEERRLLTYVNHLEAMNEDASAVAFVLOEEDRYSXGWCMPAEKTT	
		PSGKILRIIYEENDESEVEIIVTSSPMLETRADSERPKGTANRKTVEKXSEIM	
		DARGRIIDVIDKELIOPFEILFEGVEVIAAGMTPEGVAMSLILDSORTIOLVILS	
		PELFIPIEDDVMEROLIESVPDSVPLIITYESTTDWINIHDFHVFPOSHEREIEF	
		IFASBECKTFRHLKITSILKESKRYKSSGGLFAPDFKPIKEEIAITSGEMVLR	
		HGSNIQVDEVRRLVFEGETKDSPLRHLVYVSVNPDVRLTDRGYSHCISQHD	
		FFISKYNSQKRPCHVSLYKLSPEDDPTCKEFAWTLILDSAGPLDPYTPPELFSFS	
		TTGFTLYGMLYKPHDLOPKKYPVLFIYGGPQVQVNNRFKGVKFRLLNTLASLGY	
		VVVIDNRGSHRGKFGEGAFKYMGGQIEIDVQGLQYLASRYDEIDLDVGHGMSY	
		GGYLSMALMQRSDIRVATAGAPVTLIMIFYDGYRBYVNGHPDQNGCYLGSVAMQ	
		AEKPSPEPNRLILHGFLENVFAHTSILSLFLVRAGKPYDILITFOENHSIRVPS	
		GEHYELHLHYLDENLGSRTALKVTI"	
BASE COUNT	943 a 644 c 712 g 844 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	3.18e-230	Length: 3143	
Score:	2422.00	Matches: 465	
Percent Similarity:	82.30%	Conservative: 0	
Best Local Similarity:	82.30%	Mismatches: 100	
Query Match:	97.58%	Indels: 1	
DB:	6	Gaps: 1	
US-10-070-464-5 (1-465) x AX354793 (1-3143)			
Qy	1	ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu	20
Db	1180	ACAGGTACACCAATCTTAAGTCACTTTTAAGATGCAGAAATTAAGATTGAGCTGAA	1239
Qy	21	GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu	40
Db	1240	GGAAGGATCATATGATGCATAGATTAAGAACTAATCACTTTTGAGATTTCTATTGAA	1299
Qy	41	GlyValGluLysTrpIleAlaArgAlaGlyTrpThrProGluLysTyrTrpAlaTrpSerIle	60
Db	1300	GAGTTGAATATATGTCGACAGCTGATGAGTGCCTCGAGGAAATATGCTGCTCATC	1359
Qy	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle	80
Db	1360	CTACTAGATGCTCCGACACTGCCCTGCAGATAGTGTGATCTCACCCTGAATTTATTC	1419
Qy	81	ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal	100
Db	1420	CCAGTGAAGATGATGTTATGAAAGCAGACACTCATGATGACAGTGCCTGATTCTGTG	1479
Qy	101	ThrProLeuIleIleTyrGlyGluGluThrAspAlaTrpIleLeuIleHisAspIlePhe	120
Db	1480	ACGCCACTAATTTATCTATGAAGAAACACAGACATCTGATTAATATCCATGACATCTTT	1539
Qy	121	HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys	140
Db	1540	CATGTTTTTCCCAAGATCACGAAGAGAAATGAAGTTATTTTTCCTCTGAAATGCAA	1599
Qy	141	ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg	160
Db	1600	ACAGGTTTCCGTCATTATACAAATTAATCAITTTTAAAGGAAAGCAATATTAACGA	1659
Qy	161	SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla	180
Db	1660	TCCAGTGGTGGGCTGCTGCTCCAAAGTATTTCAAGTGTCTTATCAAAAGAGGATAGCA	1719

QY 181 ILeThSerGlyGluTyrPGLuValLeuGlyAArgHisGlySerAenIleGlnValAspGlu 200  
 DB 1720 ATTAACGAGTGGTGAATGGGAAGTCTTGCCGGCATGGATTATATCCAAAGTTGAGAA 1779  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeuTyr 220  
 DB 1780 GTCAAGAGCGGTGATATTTTGAAGAGCACAAAGACTCCCTTAGAGCATCACCTGTAC 1839  
 QY 221 ValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 1840 GTAGTCAGTTCGTAATCCCGAGAGGTGACAAAGCGTAGTGAACCGGTACTCTCAT 1899  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 DB 1900 TCTTGCGCATCATCAGCACTGTGACTCTTTATAGTATAGTATAGAACCAAGAT 1959  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspProThrCysLysThr 280  
 DB 1960 CCAACACTGTGTCTCTTACAAAGCTATCAAGTCTCTGAAGATGACCAACTTGCAAAACA 2019  
 QY 281 LysGluThrPheTyrPheLeuAspSerIleLeuAspSerIleProLeuProAspTyrThrProPro 300  
 DB 2020 AAGGAATTTTGGGCGACCATTTTGATTCAGACAGTCTCTTCTGACTATATCTCCCA 2079  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 2080 GAAATTTTCTCTTTGAAAGTACTAGTGGATTATGATGGAGTCTCTACAAAGCT 2139  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 2140 CATGATCTACAGCGCTGAAAGAAATATCTACTGTGCTGATATATGATGATGCTCTCAG 2199  
 QY 340 ----- 340  
 DB 2200 GTGCGATTGGTGAATAATCGGTTTAAAGAGTCAATATTCCGCTGAATACCTAGCC 2259  
 QY 340 ----- 340  
 DB 2260 TCTCTAGTATATGTTGTTAGTATAGACAAAGGGATCTCTACACGAGGCTTAAA 2319  
 QY 340 ----- 340  
 DB 2320 TTTCGAAGCGCTTAAATATATAATGGGTCAATAATGAAATTGACGATCAGGTGAAGA 2379  
 QY 340 ----- 340  
 DB 2380 CTCCAATATCTAGCTCTCGATATGATTTTCACTTGAATCGTGGGCAATCCAGCGC 2439  
 QY 340 ----- 340  
 DB 2440 TGGTCCATGAGGAGTACCTCTCCCTGATGGCATTATGACAGAGTCAGATATCTTACG 2499  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 2500 GTTGCTATTCTGGGGGCCAGTCACTCTGGATCTTATATGATATACAGATACACGGA 2559  
 QY 361 ArgTyrMetCysHisPheAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 2560 CGTTATATGGGTCACTCTGACCAAAATGAACAGGGCTATTATCTTAGAGATCTGTGGCATG 2619  
 QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuHisGlyPheLeuAsp 400  
 DB 2620 CAAGCAGAAAAGTCCCTCTGAAACCAATCTGTACTGCTTACATGGTTTCTTGAT 2679  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 DB 2680 GAAATATCCATTGTCACATACACAGTATATCTAGATTATTTTGAAGAGGCGCTGAAG 2739  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 DB 2740 CCAATATGATTTACGATCTATCTCTCAGAGAGACACAGCATAGAGTTCTCGAATCCGGA 2799

QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 DB 2800 GAACATATGAGACGATCTTTTGCATCTTACAGAAAACCTTGATCAGATTTCT 2859  
 QY 461 AlaLeuLysValIle 465  
 DB 2860 GCTCTAAAGTGATA 2874

## RESULT 15

BC040203

LOCUS BC040203 4535 bp mRNA linear PRI 03-DEC-2002

DEFINITION

Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:42570

ACCESSION

BC040203

VERSION

MGC

KEYWORDS

MGC

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 4535)

REFERENCE

1 (bases 1 to 4535)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (27-NOV-2002)

National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA.

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) &amp; Shitaki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)

Gowatne, P.H., Garcia, A.M., Lu, X., Huijck, S.W., Loulseged, H.,

Kowale, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanaavati,

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/ULNL at: <http://image.llnl.gov>

Series: IRAP plates: 71 Row: K Column: 23.

Location/Qualifiers

1..4535

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:42570 IMAGE:4824813"

/tissue\_type="Testis"

/clone\_id="NH MGC\_97"

/lab\_host="DH10B"

/note="Vector: pBluescript"

1629..4277

/codon\_start=1

/product="similar to dipeptidylpeptidase 8"

/protein\_id="AAH40203.1"

/db\_xref="GI:26007917"

/db\_xref="LocusID:54878"

/translation="MAAAETBOLGVEIFETADCEENISQDRPKLEPFYVERYSNQ

LKLADTRKTYHGMMAKAPHDPMFVKRNDPDSHDIRIYLLAMSGENRNTLFSIS

PKTINAAVLMLSMKPLDLFOATLDVGMYSREEDLREKRIGTVGASVDYHGG

TFLFOAGSGIYHVDCGPOGFTQPLRPMLVETSCNIRMDPKLCPADPMIAFIHN

DIWISNIVREERRLTYVHNELANMEDASAVAFVLOEEDRVSAGYWCRAET

PSCGATLILYEENDESEVEIITHVSPMLSTRPADSFYKCTATAPKTPKSEIMI

DAGRIIDYIDKELIOPPELIEGCVYIYRACGTPBCKAKMSILBRSQRLIVLIS

PELFIPEVDVWERQLISVPSVPLIIEETTDIWINIHIFVFPQSHEEIEF

IFASEKTEGRHLKYKTSILKEKYSRSGGLPAPDFKPIKEBIAITSGEWEVGR

HGSNIQVDEVRRLVYEGTKDSLLEHLHLYVSVYNGEVRLTRDGRYSHSCQSDHC

FISKYSNQNHPCVSLYKLSPEDDPTCKTEFMATILDSAGPLDPYTPPELFSFS



TTGFTLYGMLYKPHDLPGRKYPVPLVLYGGPOVQVNNRPKYKPRRLNTLASLVY  
 VVVIDNRGSSCHRGKFEQAFRYKMGQIEIDVOVGLDYLAIRYFIDLDRGHGWMSY  
 GGYLSMALMQRSDIPRYAIGAPVTLMIPDYDTYTERMGMHPQNRGQYVLSVAMQ  
 AEKPEPNRLLLHGFLEDEVHFAHTSILLSFLVRAGKPYDLIYPOERHSIRVPS  
 GEHEHLHLHLNLOENLSRIALAKVI"

BASE COUNT 1313 a 918 c 946 g 1358 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 5,27e-230 Length: 4535  
 Score: 2422.00 Matches: 465  
 Percent Similarity: 82.30% Conservative: 0  
 Best Local Similarity: 82.30% Mismatches: 0  
 Query Match: 97.58% Indels: 100  
 DB: 9 Gaps: 1

US-10-070-464-5 (1-465) x BC040203 (1-4535)

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 2580 ACAGGTACAGCAAAATCCTTAAAGTCACTTTAAGATGCAGAAATATGATGATGCTGAA 2639  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 2640 GGAAGATCATAGATGATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2699  
 QY 41 GlyValGluIleAlaArgAlaGlyTrpThrProGluGlyLysValAlaTrpSerIle 60  
 DB 2700 GGAGTGAATATATATGTCAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2759  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 2760 CTACTAGATGCTCTCCAGACCTCGCTACAGATAGTGTATCTCACCTGAATATATATATATC 2819  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 2820 CCAATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2879  
 QY 101 ThrProLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 120  
 DB 2880 ACGGCACATTAATCTATGAGAAACACACACATCTGATTAATATATCATGACATCTTT 2939  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 DB 2940 CATGTCTTTTCCCAAGTCACAGAGAGAAATGAGTTATTTTTCCTGCAATGCAAA 2999  
 QY 141 ThrGlyPheArgHisLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 160  
 DB 3000 ACAGGTTCCGTCATTTATACAAATTTATCATCTATTTTAAAGAAAGCAATATATACGA 3059  
 QY 161 SerSerGlyIleLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 DB 3060 TCCAGTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3119  
 QY 181 IleThrSerGlyIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 200  
 DB 3120 ATTACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3179  
 QY 201 ValArgArgLeuValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 220  
 DB 3180 GTGAGAAAGCTGATATATTTGAAAGGACCAAAAGCTCCCTTTAAGACATCACCTGTAC 3239  
 QY 221 ValValSerIleValAlaAsnProGlyValIleIleIleIleIleIleIleIleIleIleIleIle 240  
 DB 3240 GTAGTCAGTTACGTAATCTGAGAGGTGACAAAGCTGATGACCGGTGCTACTCATCAT 3299  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysIleIleIleIleIleIleIleIleIle 260  
 DB 3300 TCTTGCTGCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3359  
 QY 261 ProHisCysValSerLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 280  
 DB 3360 CCACACGTGTGTCTCTTACAGCTATCAAGCTCTGTAAGATGACCCCAATCTGCAAAACA 3419

QY 281 LysGluPheTrpAlaThrIleIleLeuAspSerAlaGlyProLeuProAspIleIleIleIleIleIleIle 300  
 DB 3420 AAGGAATTTGGGCCACCATTTTGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3479  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuIleIleIleIleIleIleIleIleIleIleIleIle 320  
 DB 3480 GAAATTTTCTCTTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3539  
 QY 321 HisAspLeuGlnProGlyLysLysValProThrValLeuPheIleIleIleIleIleIleIleIleIleIleIleIle 340  
 DB 3540 CATGATCTACAGCTGAGAAAGAAATATCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3599  
 QY 340 ----- 340  
 DB 3600 GTGCGATTGATGATTAATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTAGCC 3659  
 QY 340 ----- 340  
 DB 3660 TCTCTAGATTATGTTGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3719  
 QY 340 ----- 340  
 DB 3720 TTTGAAGCGCTTAAATATATAAATGATCAATATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3779  
 QY 340 ----- 340  
 DB 3780 CTCCAATATCTAGCTTCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3839  
 QY 340 ----- 340  
 DB 3840 TGGTCTTATGAGGATACCTCTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3899  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 3960  
 DB 3900 GTTGTCTATGCTGGGGGCCAGATCATCTGTGATCTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3959  
 QY 361 ArgIle 380  
 DB 3960 CGTTATATGAGGTACCTGACCAAGAAATGAAACAGGCTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4019  
 QY 361 GlnAlaGluIlePheProSerGluProAspArgLeuLeuLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 400  
 DB 4020 CAAAGCAAAAGTCTCCCTGAAACCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4079  
 QY 401 GluAsnValHisPheAlaHisIleThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 DB 4080 GAGAAATGCTATTTGACATACATGAT 4139  
 QY 421 ProIle 440  
 DB 4140 CCATATGATTTTACAGATCTATCTCCAGAGAGACACAGCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4199  
 QY 441 GluHisIle 460  
 DB 4200 GAAATATATGAT 4259  
 QY 461 AlaLeuLysValIle 465  
 DB 4260 GCTCTAATAAGTGATA 4274

Search completed: October 16, 2003, 00:04:57  
 Job time : 513.73 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:14:56 [Search time 341.661 Seconds  
(without alignment)]  
3673.932 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKVTKFMSRIMIDAE.....HLHYLGNGSRIALKVI 465

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+g2n.model DEV=xlh  
-Q=cg12\_1/USPTO.spool/US10070464/runat\_15102003\_113552\_24799/app\_query.faeta\_1.2652  
-DB=N-Geneseq\_19jun03 -OFMT=faeta -SUFFIX=ring -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomum2 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptlo -NOBM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10070464 @CCH 1.1 0 @runat\_15102003\_113552\_24799 -NCRP=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAFN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseq\_19Jun03: \*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and ig derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2482	100.0	1669	22 AAC85696	Nucleotide sequenc
2	2482	100.0	2668	24 ABN59775	Novel human coding
3	2482	100.0	2842	24 ABN59774	Novel human coding
4	2446.5	98.6	4523	24 ABK83325	CDNA encoding huma
5	2441.5	98.4	2161	22 AAH15009	Human cDNA sequenc
6	2422	97.6	2643	24 AAH99935	Coding sequence of
7	2422	97.6	2649	25 ABX12255	CDNA encoding huma
8	2422	97.6	2671	24 ABK83322	CDNA encoding huma
9	2422	97.6	3106	24 ABK12892	Human protease PR
10	2422	97.6	3120	22 AAC85694	Nucleotide sequenc
11	2422	97.6	3120	24 AAD89956	Human dipeptidyl p
12	2422	97.6	3143	24 AAH99934	CDNA encoding 2195
13	2418.5	97.4	4676	24 ABK83331	CDNA encoding huma
14	2402	96.8	4829	24 ABK83327	CDNA encoding huma
15	2244.5	90.4	2510	24 AAD23843	Human protease PR
16	2219	89.4	1821	24 ABV76411	Dipeptidyl peptida
17	2149	86.6	4685	24 ABK83332	CDNA encoding huma
18	1494.5	60.2	2495	24 AAD38957	Human dipeptidyl p
19	1494.5	60.2	2617	24 ABK83323	CDNA encoding huma
20	1494.5	60.2	3024	24 AAD89954	Human dipeptidyl p
21	1494.5	60.2	3716	24 ABQ75955	Human PPMW encodin
22	1494.5	60.2	4219	24 ABK83335	CDNA encoding huma
23	1494.5	60.2	4302	24 ABK83333	CDNA encoding huma
24	1488.5	60.0	3287	24 AAD38955	Alternative versio
25	1445	58.2	4180	24 ABK83339	CDNA encoding huma
26	1445	58.2	4263	24 ABK83338	CDNA encoding huma
27	1422.5	57.3	2952	24 ABK69090	CDNA encoding huma
28	1418.5	57.2	2751	24 AAD89113	Murine dipeptidyl
29	1418.5	57.2	3047	24 ABK69111	CDNA encoding huma
30	1386	55.8	4309	24 ABK83328	CDNA encoding huma
31	1377.5	55.5	1197	22 AAC85695	Nucleotide sequenc
32	1342.5	54.1	4076	24 ABK83337	CDNA encoding huma
33	1342.5	54.1	4159	24 ABK83337	CDNA encoding huma
34	1296	52.2	3262	22 AA157880	Human polynucleoti
35	1293	52.1	4037	24 ABK83341	CDNA encoding huma
36	1293	52.1	4120	24 ABK83340	CDNA encoding huma
37	1289	51.9	2982	22 AA159666	Human polynucleoti
38	1276	51.4	2801	22 AA157896	Human polynucleoti
39	1188.5	47.9	2027	21 AACT7137	Human ORFX ORF2692
40	1008.5	40.6	2461	21 AACT5835	Human ORFX ORF1390
41	860.5	34.7	1083	22 AAC85697	Nucleotide sequenc
42	845	34.0	3713	23 AB110425	Drosophila melanog
43	845	34.0	3783	23 AB106641	Drosophila melanog
44	784.5	31.6	1837	24 ABK69114	CDNA encoding huma
45	673.5	27.1	6225	23 AB110424	Drosophila melanog

## ALIGNMENTS

### RESULT 1

AAC85696  
ID AAC85696 standard; cDNA; 1669 BP.

AC AAC85696;

XX 29-JUN-2001 (first entry)

DE Nucleotide sequence encoding human DPP8 318Thr-658Val+759Ala-882Ile.

XX Human dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;

KW dipeptidyl peptidase; DPTV; T cell; cleavage; diarrhoea;

KW growth hormone deficiency; glucose level; mucosal regeneration;

KW non-insulin dependent diabetes mellitus; glucose intolerance;

XX immunosuppression; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
 FH 2..1399  
 FT CDS /\*tag= a  
 FT /product= "Human DPP8 318Thr-658Val+759Ala-882Ile"  
 PN WC200119866-A1.  
 XX  
 XX 22-MAR-2001.  
 XX  
 PF 11-SEP-2000; 2000WO-AU01085.  
 XX  
 PR 10-SEP-1999; 99AU-0002762.  
 PR 18-FEB-2000; 2000AU-0005709.  
 XX  
 PA (UNSY ) UNIV SYDNEY.  
 XX  
 PI Abbott CA, Gorell MD;  
 XX  
 DR WPI; 2001-281520/29.  
 DR P-PSDB; AAB47189.  
 XX  
 PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving  
 PT substrates, identifying inhibitors of DPP8 catalytic activity which  
 PT have therapeutic uses, and for detecting activated T cells  
 XX  
 PS Claim 21; Page 72-74; 78pp; English.  
 XX  
 CC The sequences given in AAC05695-97 encode fragments of human  
 CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for  
 CC H-Ala-Pro-pNa, H-Gly-Pro-pNa and H-Arg-Pro-pNa. Therefore, it is a  
 CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable  
 CC of hydrolysing the peptide bond C-terminal to Pro in each of these  
 CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for  
 CC cleaving a substrate, and for detecting an activated T cell which  
 CC involves measuring the level of DPP8 gene expression in a T cell. The  
 CC level of DPP8 expression is detected by detecting the amount of DPP8  
 CC RNA in the cell. It is also useful for identifying a molecule capable  
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules  
 CC identified as inhibiting DPP8 catalytic activity may be useful for  
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels  
 CC in non-insulin dependent diabetes mellitus and other disorders  
 CC involving glucose intolerance, enhancing mucosal regeneration and  
 CC as immunosuppressants.  
 XX  
 SQ Sequence 1669 BP; 524 A; 336 C; 336 G; 473 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.13e-254 Length: 1669  
 Score: 2482.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0  
 US-10-070-464-5 (1-465) x AAC05696 (1-1669)  
 QY 1 ThrGlyThraAlaenProLySValThrPheLysMetSerGluIleMetIleASPAlaGlu 20  
 DB 2 ACAGGTACAGCAAACTCTTAAGTCACTTTTAAGATGTCAGAAATTAATGATTGAGTCTGAA 61  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 62 GGAAGGATCATAGATGTCATAGATTAAGAACTTAATCACTTTGAGATTCTATTGGAA 121  
 QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 DB 122 GGAATTGATATATTGCGACAGCTGAGTGAAGTCTCTGAGGAAATATGCTGATCCATC 181  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 182 CTACTAGATCGCTCCCAAGACTCGCTCAAGATAGTGTATCTCACTGAATTATTTATC 241

QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 242 CCAGTAGAGATGATGTTATGAGAAAGCAGAGACTCATTTGATGACGCTGATTCGTG 301  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 302 ACCGCTAATTTATCTATGAGAAACACAGACCTTGATTAATATCCATGACATCTTT 361  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 DB 362 CATGTTTTTCCCAAGATCAAGAAAGGAATGAGTTATTTTGGCTCTGAAATGCAAA 421  
 QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 DB 422 ACAGTTTCCGTCATTTATCAAAATTAACATCTATTTTAAAGGAAACAAATATTAACGA 481  
 QY 161 SerSerGlyGluLeuProAlaProSerAspPheLysCysProIleLysGluIleAla 180  
 DB 482 TCAGTGTGGCTGCTGCTCCAGATGATTTCAAGTGTCTTATCAAGAGAGATAGCA 541  
 QY 181 IleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 DB 542 ATTAACAGTGTGATGAGAAAGTTCTTGCCCGCATGATCTATATCCAAAGTGAAGA 601  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeuTyr 220  
 DB 602 GTCAAGAGCTGTGATTTTGAAGGCACCAAGACTCCCTTTAGAGCATCACCTGTAC 661  
 QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 662 GTATCTCAATTAAGTAATTCCTGGAGAGTGAAGGCTGACTGCCGGCTACTCATCAT 721  
 QY 241 SerCysCysIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGlnLysAsn 260  
 DB 722 TCTTGTCGATCATGTCAGCATGCTGACTTCTTTTAAGTATGATTAACAGAGAT 781  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 782 CCACACGTGTGTCCCTTTCAMAGCTTCAAGTCTTAAGATGACCAACTTGCMAAACA 841  
 QY 281 LysGluPheThrPalThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 DB 842 AAGGAATTTTGGCCACCATTTTGGATTTCAGCAGGTCCTCTTCTGACTATATCTCTCA 901  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 902 GAATTTTCTCTTTGAAGATCACTGATTTATCATTTGATGAGTCTCTACAAAGCT 961  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 962 CATGATCTACAGCTGGAAGAAATATCTCACTGTCTGTTCATATATGGTGTCTCCAG 1021  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 1022 GTTCTATTGCTGGGCCCCAGTCACTCTGTGATCTTCTATGATACAGATACAGGAA 1081  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 1082 CGTATATGGGTACCCCTGACACAGATGAACAGGCTATTATCAAGATCTGTGGCAGT 1141  
 QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 DB 1142 CAACGACAAAGATTCCTCTGAAACAAATCGTTATCTGCTTCAATGATGTTCTCGAT 1201  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 DB 1202 GAGATATGTCATTTTGCACATACAGATATTAATGACTTTTAAAGAGGCTGGAAG 1261  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 DB 1262 CCATATGATTTTACAGATCTATCTCCAGAGACACACAGCATTAAGAGTTCCGTAATCGGA 1321  
 QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460

Db 1322 GAACATTATGAAGTGCATCTTTGGCACTCACTTCAAGAAAACCTTGATCAGTATTTGCT 1381  
QY 461 A1a1eUlyValIle 465  
Db 1382 GCTCTAAAAGTGATA.1396

## RESULT 2

ABNS9775

ID ABNS9775 standard; cDNA; 2668 BP.

AC ABNS9775;

DT 28-JUN-2002 (first entry)

DE Novel human coding sequence SEQ ID NO: 186.

Human; antianemic; vulnerary; antiinflammatory; immunomodulator;  
antifertility; cerebroprotective; cyostatic; rheumatic; gene therapy;  
neuroprotective; antiparkinsonian; protein therapy; Estf;  
expressed sequence tag; gene; ss.

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Dymanac RT;

DR WPI; 2002-292408/33.

DR P-PSDB; ABB97362.

PT An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -

PS Claim 1; SEQ ID NO 186; 509pp; English.

CC The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a coding sequence of the  
CC invention.

SQ Sequence 2668 BP; 796 A; 564 C; 592 G; 716 T; 0 other;

## Alignment Scores:

Pred. No.: 2,3e-254 Length: 2668  
Score: 2482.00 Matches: 465  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0

US-10-070-464-5 (1-465) x ABNS9775 (1-2668)

QY 1 ThrglyThrAlaAspProLysValThrPhelyMetSerGluIleMetIleAspAlaGlu 20  
Db 1011 ACAAGTACAGCAAAATCCTAAGTCACTTTAAGATGTGAGAAATATGATGATGCTGAA 1070

QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
Db 1071 GGAAGATCATAGATTCATAGTAAAGAACTAATTCACCTTTAGATTCTAATTGAA 1130  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
Db 1131 GGAGTTGAATATATTCAGAGCTGGATGAGTCTCCTGAGGAAATATGCTTGTCATC 1190  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1191 CTAAGTATGATGCTCCAGACCTGCTTACAGATAGTGTATGATCCACCTGATATTATTC 1250  
QY 81 ProValGluAspAspValMetGluArgGluArgLeuIleGluSerValProAspSerVal 100  
Db 1251 CAGTAGAAGATGATGCTTATGAAAGGACAGACATCATTAAGTCACTGCTGATTCGTG 1310  
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
Db 1311 ACGCCACTAATATATCATGAGAAACACAGCATCTGGATTAATATCATGATCATCTTT 1370  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
Db 1371 CATGTTTTTCCCAAGTCACAGAGAGAAATTTGATTATTTTGGCTCTGAAATGCAAA 1430  
QY 141 ThrglyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
Db 1431 ACAAGTTTCGATCATTTTATACAAATTCATCATTTTAAAGAAAGCAAAATTAACGA 1490  
QY 161 SerSgrGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
Db 1491 TCCAGTGGGCTGCTGCTGCTCAAGTATTTCAAGTGTCCCTATCAAGAGAGATAGCA 1550  
QY 181 ILeThrSerGlyGluTyrPglValIleGluLysArgHisGlySerAsnIleGlnValAspGlu 200  
Db 1551 ATTACAGGTGGTGAATGGAGAGTTCTTCCGCGCATGATCAATATCAAGATTGATGAA 1610  
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLysTyr 220  
Db 1611 GTCAAGAGCTGTATATTTTGAAGCACCAAGACTCCCTTTAGAGATCACTGTAC 1670  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgLysTyrSerHis 240  
Db 1671 GTAAGTCACTTACGTAAATCCCTGAGAGAGTCAAGAGTCACTGCTGCTACAT 1730  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
Db 1731 TCTTCTCATCATGACACACACTGATCTTTATTAAGTATAGTAAACAGAAAGAT 1790  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
Db 1791 CCACCTGTGTCTCCCTTACAGCTATCAAGTCTCTGAAGAGACCCCACTTGCAAAACA 1850  
QY 281 LysGluPheThrPalaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
Db 1851 AAGGAATTTTGGGCCACATTTTGGATTCACAGAGTCTCTTCCGACTATATCTCTCCA 1910  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
Db 1911 GAAATTTTCTTTTGAAGAAGTACTAGATTATGATGATGAGATGCTCTCAAGCCT 1970  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValIleLeuPheIleTyrGlyGlyProGln 340  
Db 1971 CATGATCTACAGCCCTGGAAGAAATATCTCTGCTGCTTCAATATATGATGCTCTAG 2030  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlu 360  
Db 2031 GTTGCTATTTGCTGGGGGCCAGTCACTGTGGATCTTTCTATGATACGATACGGA 2090  
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
Db 2091 CGTTATATGGGTCAACCTCGAACCAAGATGAACAGGGCTATTCTTAGATCTGTGGCCATG 2150  
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400

Db 2151 CAAGCAGAAAGTCCCTCTGAACCAATCGTTTACTGCTTACATGGTTCCGGAT 2210  
 Qy 401 GUAaValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 Db 2211 GAGAACTCTCATTTTCCACATACCAGATATTTACTGATTTTAAAGGGCTGGAAAG 2270  
 Qy 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlnSerGly 440  
 Db 2271 CCATATGATTTTACAGATCTATCTCTCAAGAGACACAGCATTAAGATTCCTGAATCGGGA 2330  
 Qy 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluSerLeuGlySerArgIleAla 460  
 Db 2331 GAACATTAATGACAGCTGCACTTTTGACACTTCAAGAAACCTTGATCAGTATTGCT 2390  
 Qy 461 AlaLeuLysValIle 465  
 Db 2391 GCTCTAAAGTGATA 2405

RESULT 3  
 ABNS9774 ID ABNS9774 standard; cDNA; 2842 BP.  
 AC ABNS9774;

DT 28-JUN-2002 (first entry)  
 DE Novel human coding sequence SEQ ID NO: 185.

Human; antinaeamic; vulnereary; antiinflammatory; immunomodulator;  
 antiinfertility; cerebroprotective; cytoskeletal; rheumatic; gene therapy;  
 neuroprotective; antiparkinsonian; protein therapy; EST; expressed sequence tag; gene; ss.

OS Homo sapiens.  
 PN MO200222660-A2.  
 PD 21-MAR-2002.  
 PF 10-SEP-2001; 2001WO-US26015.  
 PR 11-SEP-2000; 2000US-0659671.

PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QH, Ren F;  
 PI Xue AJ, Yang Y, Weinman T, Drmanac RT;  
 DR MPI: 2002-292408/33.  
 DR P-PSDB; ABB97361.

PT An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis -  
 PS Claim 1; SEQ ID NO 185; 509pp; English.

CC The present invention provides the protein and coding sequences of 444  
 CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to regulate  
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a coding sequence of the  
 CC invention.

SQ Sequence 2842 BP; 857 A; 592 C; 635 G; 758 T; 0 other;

Alignment Scores: 2.53e-254 Length: 2842  
 Pred. No.:

Score: 2482.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 24 Gaps: 0

US-10-070-464-5 (1-465) x ABNS9774 (1-2842)

Qy 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 1185 ACAGATACAGCAAAATCTTAAGTCATTTTAAGATCTCGAAATATATGATTGATCGTGA 1244  
 Qy 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 1245 GGAAGATCATGATGATCTCATAGTAAAGAACTAATCAACTTTTACAGATTCTATTGAA 1304  
 Qy 41 GlyValGluTyrIleLeuAlaArgAlaGlyTyrPheThrProGlnGluLysTyrIleAspSerIle 60  
 Db 1305 GAGATTGAATATATATGCTCAAGCTGATGACTCTCGAAGGAAATATGCTTGCTCATC 1364  
 Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1365 CTACTAGATCGCTCCAGACTCGCTCAGATAGTGTGATCTCAGCTGAATATTATTATC 1424  
 Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1425 CCAGTGAAGATGATGTTTATGAAAAGGCAGACATCATTAAGTCACTGCTGATTCGTG 1484  
 Qy 101 ThrProLeuIleIleTyrGluGluThrPheAspIleThrIleAsnIleHisAspIlePhe 120  
 Db 1485 AGCCACTAATATCTATATGAAGAACACAGACATCTGATTAATATCATGACATCTTT 1544  
 Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 Db 1545 CATGTTTTCCCAAGATCAGCAAGAGAAATGAGTTATTATTTGCCCTCGAATCGAAT 1604  
 Qy 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 Db 1605 ACAGATTTCCGTCATTTATTAACAAATATACATCTATTATTAAGCAAAATATTAACGA 1664  
 Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 Db 1665 TCCAGTGTGGGCTGCTCTCCAGATGATTTCCAGTGTCTTAACAAAGAGATGAGCA 1724  
 Qy 181 IleThrSerGlyGluTyrPheGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 Db 1725 ATTACAGAGGAGATGGAAGATTCTTGCCCGGAGATGATCTTAATATCCAAGTTGATGA 1784  
 Qy 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220  
 Db 1785 GTCAAGAGGCTGTATATTTTGAAGGCACCAAAAGCTCCCTTTAGAGCATCACCTGTAC 1844  
 Qy 221 ValIleSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1845 GTAGTACGTTACGTAAATCTCTGAAGGTGACAAAGGCTACGACCGTGCTACTACAT 1904  
 Qy 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 Db 1905 TCTTGCTGATCAGTCAGACAGTGAATCTTTATTAAGTAATATATGTAACCAAGAAAT 1964  
 Qy 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 Db 1965 CCACACTGTGTCTCTTTACAAAGTATCAACTCTGAAGAGAGACCAACTTGCAAAACA 2024  
 Qy 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 2025 AAGGAATTTGGGCCACATTTTGAATTCAGAGATCTCTTCCGACATATCTCTCCA 2084  
 Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 Db 2085 GAATTTTCTCTTTGAAGATGACTAGATTTATCATGTATGAGATGCTCAAGCCT 2144  
 Qy 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGln 340

Db	2145	CATGATCTACAGCCCGGAAAGAAATATCTACTAGCTGCTTCATATAGTGTCCTCAG	2204
Oy	341	ValAlaIlealaGlyAlaProValThyLeuTrpIlePheTyraaPthGlyTyrtHcIu	360
Db	2205	GTTGCTATATGCTGGGGCCCCAGATCACTCTGTGGATCTTCTATGATACAGGATACACGAA	2264
Oy	361	ArgTyrmctGlyHisProAspGlnasnGlnGlnGlyTyrtTyrtLeuGlySerValAlaMet	380
Db	2265	CGTATATAGGCTCACCTCCAGTACCCAGNATGACAGGGCTATTTACTTATAGATCTGTGGCCATG	2324
Oy	381	GlnAlaGlnTyrsPheProSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAsp	400
Db	2325	CAAGCAGAGAAAGTTGCTCCCTCTGAAACCAATGCTTATCGTCTTACATAGGTTTCCGTGAT	2384
Oy	401	GluAsnValAlaPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys	420
Db	2395	GAGAAATGCTCATTTTGACATATCCAGATATATTACGAATTTTATATGATAGGGCTGGAAAG	2444
Oy	421	ProTyraaPheLeuGlnIleTyrtProGlnIuArgHisSerIleArgValProGlnIuSerGly	440
Db	2445	CCATATGATTTTACAGATCTATCTCTCAGAGAGACACACCATATAGAGTTCTCGAATCCGGA	2504
Oy	441	GluHisTyrtGlyIuLeuHisIleuLeuHisTyrtTyrtLeuGlnIuAsnLeuGlySerArgIleAla	460
Db	2505	GAACTATATAGAACTGCATCTTTTGGACATACCTTCAAGAAAACCTTGGATACAGTATGCT	2564

CC The present invention relates to the isolation of novel human serine  
CC processes referred to as dipeptidyl peptidase IV (DPPV)-related  
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)  
CC and nucleic acids encoding them are useful for treating infections  
CC such as fungal, bacterial, protozoan and viral infections, particular  
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, osteoporosis, angina pectoris,  
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
CC schizophrenia), and dyskinesias. These may also be used in discovering  
CC therapeutic agents for the treatment of reproductive, inflammatory and  
CC metabolic disorders. ABRK3322-ABK3343 encode human DPRP proteins.  
XX

Sequence 4523 BP; 1384 A; 828 C; 940 G; 1371 T; 0 other;

**Alignment Scores:**

Pred. No.:	0	3,16e-250	Length:	4522
Score:		2446.50	Matches:	463
Percent Similarity:		99.57%	Conservative:	0
Best Local Similarity:		98.57%	Mismatches:	1
Query Match:		99.57%	Indels:	2
DB:		24	Gaps:	1

US-10-070-464-5 (1-465) X ABK83325 (1-4523)

Qy	461	AlaLeuLysValIle	465
Qy	461	AlaLeuLysValIle	465
Db	2565	GCTCTAAAGTGATA	2579
Qy	1	ThrGlyThrAlaAnProLysValThrPhenylsMetSerGluIleMetIleAspAlaGlu	20
Db	1165	ACAGGTACAGCAAAATCCAAAGTACATTTAAAGTGCACAAATAAAGATTCAGCTGAA	1224

ID ABK83325 standard; cDNA; 4523 BP.

DT 12-AUG-2002 (first entry)

DE CDNA encoding human DPRP-1 splice variant #1.

KM Human, serine protease; dipeptidyl peptidase IV-related protein; DPP-  
 KM DPPIV, infection; human immunodeficiency virus; HIV-1, HIV-2; pain;  
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KM dyskinnesia; reproductive disorder; inflammatory disorder;  
 KM metabolic disorder; gene; ss.

OS Homo sapiens.

PN WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR ) FERRING BV.

PI Qi S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI; 2002-444178/47.

XX

PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and

Pr

PS Disclosure; Page 61-62; 113pp; English.

X

QY	1	ThrlGlyThrAlaAspProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu	20
DB	1165	ACAGGTACACCAATCTTAAGTACCTTTTAAGATGCACAGAAATAAAGATTGACCTGCA	1224
QY	21	GIYArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu	40
DB	1225	GGAAGGATCATAGATGTCATAGATTAAGAACTAATTCACCTTTTGAGATTCATTTGAA	1284
QY	41	GIYValGluIlyrIleAlaArgAlaGlyTrpThrProGluIlyrIleAlaTrpSerIle	60
DB	1285	GGATTTAAATATATGTCACAGCTGGATGAGATCTCTGAGGAAATATGCTGGTTCATC	1344
QY	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle	80
DB	1345	CTACTAATTCGCTCCAGATCTCGCTCAAGATAGTGTGATCTCACTGAATTTATATC	1404
QY	81	ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal	100
DB	1405	CCAATAGAAAGATGATGTTATGAAAGCAGAGACTCATTTGATGATGCTGATTCGTG	1464
QY	101	ThrProLeuIleIleIleYrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePhe	120
DB	1465	ACGGCACTAATTAATCTATGAAAGAAACAACAGACTTCGATTAATATCCATGACATCTTT	1524
QY	121	HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys	140
DB	1525	CATGTTTTTCCCAAAATCCAGAAAGCAAGAAATAGATTTATTTTGGCTCTGAATGCMA	1584
QY	141	ThrGlyPheArgHisLeuYrIlySerIleThrSerIleLeuLysGluSerIlyrIlyAsArg	160
DB	1585	ACAGGTTTCGTCATTTATACAAATATACATCTATTTTAAAGGAAACCAATATTAACGA	1644
QY	161	SerSerGlyIlyLeuProAlaProSerAspPheLysCysProIleLysGluIleAla	180
DB	1645	TCCAGTGTGGGCTGCTGCTCCAAAGATTTCAAGTGTCTTCAAAAGAGGAGATGCA	1704
QY	181	IleThrSerGlyGluTrpGluValIleLysIlyArgHisGlySerAsnIleGlnValAspGlu	200
DB	1705	ATTAACAGATGTGATGGGAAGTCTTGCGCGGATGATGATTAATATCCAAAGTTGAAGA	1764
QY	201	ValArgArgLeuValIlyrPheGluGlyThrLysAspSerProLeuGluIlyHisIleLeuYr	220
DB	1765	GTCGAAGGCTGGATATTTTGAAGGCAACAAAGCTCCCTTTAAGGATCACTGCTAC	1824
QY	221	ValValSerIlyrValAsnProGlyGluValThrArgLeuThrAspArgIlyrSerHis	240

Db 1825 GTAGTCAGTACGTAATTCCTGGAGAGGTGACAAAGCTGACTGACCGTGCTACTACAT 1884  
 Qy 241 SerCysCysIleSerGlnHisCysaspPheIleSerIleTySerAsnGlnIleAsn 260  
 Db 1885 TCTTGCTGCATCAGTCACACACCTGTCATCTTTATTAAGTAAAGTAAAGTAAAGTAA 1944  
 Qy 261 ProHisCysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIleThr 280  
 Db 1945 CCACAGCTGTGTCTCCCTTACAGATCAAGTCTCTGAAGATGAGCCAACTGGCAAAACA 2004  
 Qy 281 LysGluPheThrPalaThrIleLeuaspSerIleagIlePoleuProaspTyrThrProPro 300  
 Db 2005 AAGGAATTTGGGCCACCATTTGGATTGAGAGGTCTCTTCGACTATACCTCTCCA 2064  
 Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyIleuLeuTyrIleAsp 320  
 Db 2065 GAAATTTTCTCTTTGAAAGTAACTACTGATTTATGATTTAGGATGCTCTACAAAGCT 2124  
 Qy 321 HisaspLeuGlnProGlyIleIleTyIleTyProThrValIleuPheIleTyIleGlyProGln 340  
 Db 2125 CATGATCTACAGCTCGGAAAGAAATATCTACTGCTGCTTCAATATGCTGCTCG--- 2180  
 Qy 341 ValalaIleIleagIleValProValThrIleuThrIlePheTyrAspThrGlyTyrThrGlu 360  
 Db 2181 GTGCTATGCTGGGGCCCCCAGTCACTCTGTGATCTTCTATGATACAGATACCGGAA 2240  
 Qy 361 ArgTyrMetGlyHisProAspGlnaspGlnaspGlnaspGlnaspGlnaspGlnaspGln 380  
 Db 2241 CGTTATATGGGTCAACCCCTGACACAGAAATGAAACAGGCTATTTAGATGCTGTGGCCATG 2300  
 Qy 381 GlnaspGlyIlePheProSerGluProaspArgLeuLeuLeuLeuHisGlyPheLeuAsp 400  
 Db 2301 CAACAGAAAGTTCCCTCTGAAACCAATGTTTACTGCTTCAATGATGCTTCTCTGAT 2360  
 Qy 401 GluaspValHisPhealaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyAsp 420  
 Db 2361 GAGAAATGCTCATTTTGGACATACCGATATATTACTGAGTTTATGAGGCTGGAAAG 2420  
 Qy 421 ProTyrAspLeuGlnIleTyrProGlnIleuArgHisSerIleArgValProGluSerGly 440  
 Db 2421 CCATATGATTTACAGATCTATCTCTCAGAGAGACACACATTAAGGTTCCGTAATCGGGA 2480  
 Qy 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnIleuLeuGlySerArgIleAla 460  
 Db 2481 GAAATATTAGAACGTCATCTTTTGCATACCTTCAAGAAACCTTGATCAAGTATTCCT 2540  
 Qy 461 AlaLeuIleValIle 465  
 Db 2541 GCTCTAAAGTGATR 2555  
 RESULT 5  
 AAH15009  
 ID AAH15009 standard; cDNA; 2161 BP.  
 XX  
 AC AAH15009;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:12963.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI, 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 12963; 2537bp + CD ROM; English.  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to a  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and  
 CC AAH1633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 2161 BP; 662 A; 429 C; 452 G; 618 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 3,54e-250 Length: 2161  
 Score: 2441.50 Matches: 464  
 Percent Similarity: 90.27% Conservative: 0  
 Best Local Similarity: 90.27% Mismatches: 1  
 Query Match: 98.37% Indels: 49  
 DB: 22 Gaps: 1  
 US-10-070-464-5 (1-465) X AAH15009 (1-2161)  
 Qy 1 ThrGlyThrAlaAsnProIleValThrPheIleMetSerGluIleMetIleAspAlaGlu 20  
 Db 370 ACAGGTACAGCAAAATCCTAAGTCACTTTAAAGATGTCAGAAATTAAGATGATGCGGAA 429  
 Qy 21 GlyArgIleIleAspValIleAspIleGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 430 GGAAGCATCATAGATGTCATAGTAAGAACTAATTCAACCTTTTGAGATTCATTGGAA 489  
 Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyIleTyrIleTyrSerIle 60  
 Db 490 GGAAGTGAATTAATTTGCGACAGAGCTGATGAGCTCTGAGGGAATATGCTGTGCATC 549  
 Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 550 CTACTTAATGCTCCCGACTCGCTTACAGATGCTTGAATCTCACTGAAATTAATTAATC 609  
 Qy 81 ProValGluAspAspValIleMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 610 CCACTAGAGATGATGTTATGAAAGGCAAGACTCAATTAGTACAGTCCGCTGATTCGTG 669  
 Qy 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleLeuIleHisAspIlePhe 120

```

Db      670  ACCGCAATTAATCTATGAAAGAAACAAGACATCGATTAATATCCATGACATCTTT 729
Qy      121  HIsValPheProGlnSerHisgluGluGluileGluPheIlePheAlaSerGluCysLys 140
Db      730  CATGTTTTTCCCAAGTACAGAGAGAAATTGAGTTATTTTGGCTCGAATGCAAA 789
Qy      141  ThGlyPheArgHisLeuTyrlsLeuIleThSerIleLeuLysGlnSerLysTyrlsValArg 160
Db      790  ACGAGTTTCCGCTCTTATACAAAATTACATCTTTTAAAGAAAGCAAAATTAAGCA 849
Qy      161  SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysgluGluIleAla 180
Db      850  TCCAGTGTGGGCTGCTGCTGCCAAGATTTCAAGGTCTCTATCAAAAGAGATATGCA 909
Qy      181  IleThSerGlyGluTTPGluValLeuGlylAArgHisglYserAnIleGlnValAspGlu 200
Db      910  ATTACCGATGTGATGAGGAAGTCTTGGCCGATGATCTAATATCCAAAGTTGATGAA 969
Qy      201  ValArgArgLeuValTyrlsPheGluGlyThrlsAspSerProLeuGlnHisIleLeuTyrls 220
Db      970  GTCCAGAGGCTGTATATTTTGAAGGACCAAGACTCCCTTGAAGCATCACTGTAC 1029
Qy      221  ValValSerTyrlsValAsnProGlyGlyValThArgLeuThrAspArgGlyTyrlsSerHis 240
Db      1030  GTAGTCAGTACGTAATCTCGAGAGGTGACAAAGGCTGACCGTAGCTACCAT 1089
Qy      241  SerCysCysIleSerGlnHisCysAspPheIleSerIleSerIleSerIleValAsn 260
Db      1090  TCTTGCTGCATCGATCGACACTGTGACTTTTAAAGTAAGTAACTAACCGAAGAAAT 1149
Qy      261  ProHisCysValSerLeuTyrlsLysLeuSerSerProGluAspAspProThrCysLysThr 280
Db      1150  CCACACTGTGTGCTCTTGAAGCTATCAAGCTATCAAGTCTTGAAGTACCCAACTGCAAAACA 1209
Qy      281  LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrlsProPro 300
Db      1210  AAGGAATTTGGGCGCACCATTTTGGATTTCAGAGGTCTCTCTGACTATGACTCTCCCA 1269
Qy      301  GluIlePheSerPheGlnSerThrThrGlyPheThrLeuTyrlsGlyMetLeuTyrlsProPro 320
Db      1270  GAAATTTTCTCTTTGAAATGACTACTGATTTACATTTGATGGAGTCTCTCAAGACCT 1329
Qy      321  HisAspLeuGlnProGlyLysLysTyrlsProThrValLeuPheIleTyrlsGlyProGln 340
Db      1330  CAGGATCTACAGGCTGGAAGAAATATCTACTGTGTCTTCAATATGTTGTTCTCTCAG 1389
Qy      340  ----- 340
Db      1390  GTGCAGTTGGTGAATATCGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCCCTAAGC 1449
Qy      340  ----- 340
Db      1450  TCTCTAGGTTATGTGTTAGTAGATAGACACAGGGGATCTGTCAACCGAGGGCTTAA 1509
Qy      341  -----ValAlaIleAlaGlyAlaProValThrLeuTyr 351
Db      1510  TTTGAAGGGCGCTTAAATATATAATATGTTGCTATTCCTGGGCGCCCAAGTCACTCTGTGG 1569
Qy      352  IlePheTyrlsAspThrGlyTyrlsGluArgTyrlsMetGlyHisProAspGlnAsnGlnGln 371
Db      1570  ATCTTCTATGATACAGGATACAGGAACGTTATATGGGTACACCTGACCGAATGAACAG 1629
Qy      372  GlyTyrlsTyrlsLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArg 391
Db      1630  GGCTATTACTTACGATCTGTGGCAATGCAAGCAAGAAAGTTCCCTCTGAACCAAAATCGT 1689
Qy      392  LeuLeuLeuLeuHisglYlPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 411
Db      1690  TTAAGTCTCTTACATGAGTTTCCGGAATGAGAATGTCATTTTGCAATACACGATATATTA 1749
Qy      412  LeuSerPheLeuValArgAlaGlyLysProTyrlsAspLeuGlnIleTyrlsProGlnIuArg 431

```

```

Db      1750  CTGAGTTTTTAGTAGAGGCTGGAAAGCATATGATTTACATCTATCTCAGAGAGA 1809
Qy      432  HisSerIleArgValPProGlnSerGlyGluHisTyrlsGluLeuHisLeuHisTyrlsLeu 451
Db      1810  CACAGCATTAAGAGTCTCGATCGGAGAACATTATTAACATGCAATCTTTTGCACTACTT 1869
Qy      452  GlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 465
Db      1870  CAGGAAAACCTTGATCAGTATTTGCTCTCTAAAGATGA 1911

RESULT 6
AAH99935
ID  AAH99935 standard; cDNA; 2643 BP.
XX
AC  AAH99935;
XX
DT  12-APR-2002 (first entry)
XX
DE  Coding sequence of 21953 human prolly1 oligopeptidase.
KM  21953 prolly1 oligopeptidase; antibody; proline; endopeptidase;
KM  cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KM  neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KM  antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KM  diabetes mellitus; arthritis; multiple sclerosis; asthma;
KM  Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
OS  Homo sapiens.
XX
PN  M0200179473-A2.
XX
PD  25-OCT-2001.
XX
PF  11-APR-2001; 2001M0-US40483.
XX
PR  18-APR-2000; 2000US-197508P.
XX
PA  (MILL-) MILLENNIUM PHARM INC.
XX
PT  Meyers RA, Williamson M;
XX
DR  WPI; 2002-034353/04.
XX
P-PSDB; AAG78415.
XX
PT  New polypeptides 21953, member of human prolly1 oligopeptidase family,
PT  useful as diagnostic targets and therapeutic agents for controlling
PT  cancer, lymphoma and leukemia -
XX
PS  Claim 2; Page; 121pp; English.
XX
CC  This invention relates to an isolated 21953 human prolly1
CC  oligopeptidase. Which is cytosolic, antidiabetic, antiarthritic,
CC  neuroprotective, antithyroid, dermatological, antipapillary,
CC  antiasthmatic, ophthalmological, antiinflammatory, neotropic,
CC  antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC  antineural, cardiant, antihypertensive, anorectic and
CC  metabolic in its action. Uses include gene therapy, expression or
CC  activity of 21953 protein modulator, it is useful for identifying a
CC  compound which binds to it and can be used in preventing, treating
CC  or detecting a cellular proliferative or differentiative disorder.
CC  The 21953 molecules can act as novel diagnostic targets and therapeutic
CC  agents for controlling disorders associated with the aberrant activity
CC  or degradation of peptide hormones e.g., disorders associated with cell
CC  differentiation and proliferation such as cancer, immune function,
CC  reproductive, neurological and cardiovascular function. The 21953
CC  molecules are thus useful for treating and preventing cellular
CC  proliferative and differentiative disorders, haematopoietic neoplastic
CC  disorders, immune disorders such as autoimmune diseases, diabetes
CC  mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC  neuronal disorders, demyelinating diseases, vascular diseases and
CC  metabolism or pain disorders. This sequence represents the cDNA
CC  encoding sequence of 21953 human prolly1 oligopeptidase. This
CC  sequence represents the coding sequence of 21953, being the sequence

```

CC in between the start and the stop codon of the sequence represented in  
 CC AAH99934.

Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;

# Alignment Scores:

Pred. No.:	5.8e-248	Length:	2643
Score:	2422.00	Matches:	465
Percent Similarity:	82.30%	Conservative:	0
Best Local Similarity:	82.30%	Mismatches:	0
Query Match:	97.58%	Indels:	100
DB:	24	Gaps:	1

US-10-070-464-5 (1-465) x AAH99935 (1-2643)

QY	1	ThGlyThrAlaenProLyValThrPhelyMetSerGluIleMetIleAspAlaGlu	20
DB	949	ACAGGTACAGAAATCTTAAGTCACTTTTAAGTCAAGAAATTAAGTGAAGCTGAA	1008
QY	21	GIYAAGIleIleAspValIleAspLyGluLeuIleGlnProPheGluIleLeuPheGlu	40
DB	1009	GGAAGATCATAGATGATGATAGAGAACTAATCACTTTTGAGATTCTATTGAA	1068
QY	41	GIValaGluTyrIleAlaArgAlaGlyTyrThrProGluGluValTyrAlaTyrSerIle	60
DB	1069	GGAATTAATATATATGCGAGAGCTGAGTGAATCTCTAAGGAAATATGCTGTGATC	1128
QY	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleLeuIleSerProGluLeuPheIle	80
DB	1129	CTACTAATATGCTCCAGAGCTGCTGAGTACTGATGATCTCACTGAATTTATTATC	1188
QY	81	ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal	100
DB	1189	CCAGTAAAGATGATGTTATGAAAGGACAGAGCTCATGTCAGTCCGATTCGTG	1248
QY	101	ThrProLeuIleIleTyrGluGluThrThraSpleIleTyrIleAsnIleHisAspIlePhe	120
DB	1249	ACGCACTAATATCTATGAAAGAAACAACAGACATCGATTAATATCCATGACATCTTT	1308
QY	121	HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeys	140
DB	1309	CAGTCTTTTCCCAAGTCAAGAGAGAAATGATTTATTTTGGCTCTGAATGCAAA	1368
QY	141	ThGlyPheArgHisLeuTyrIleThrSerIleLeuLyGluSerIleTyrIleArg	160
DB	1369	ACAGGTTCCGCTATTATACAAATTAACATCTATTAAAGGAAACCAATATTAACGA	1428
QY	161	SerSerGlyGlyLeuProAlaProSerAspPheIleCysProIleLysGluGluIleAla	180
DB	1429	TCCAGTGGTGGCTGCTGCTCCCAAGTATTTCAAGTCTCTATCAAGAGAGATGCA	1488
QY	181	IleThrSerGlyGluTyrGluValIleGluIleArgHisGlySerAsnIleGlnValAspGlu	200
DB	1489	ATTACAGTGGTGAATGGAGAGTCTTGCCCGGATGATCTAATATCCAAAGTTGATGAA	1548
QY	201	ValArgArgLeuValTyrPheGluGlyThrIleAspSerProLeuGluHisIleTyr	220
DB	1549	GTCAGAGGCTGGTATATTTTGAAGGCAACAAAGACCTCCCTTGAAGCATCACTGTAC	1608
QY	221	ValIleSerTyrValIleAsnProGlyGluValIleThraGluThrAspArgGlyTyrSerHis	240
DB	1609	GTAAGTACGTAATCTGAGAGGTACCAAGGCTGACCTGACCTGCTACTCAT	1668
QY	241	SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnIleAsn	260
DB	1669	TCTTGCTGATCAGTCACTGACCTGATCTCTTATAGTAAATAGTATACAGAAAGAT	1728
QY	261	ProHisSerValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIleThr	280
DB	1729	CCACACGTGTGCTCTTACAAAGCTATCAAGTCCGAAATATACCCAACTTCGAAACA	1788
QY	281	LyGluIleThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro	300

DB	1789	AAGAAATTTGGGCCACCAATTTGGATTGACAGAGCTCTTCTCTGATCTACTCTCA	1848
QY	301	GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIlePro	320
DB	1849	GAATTTTCTCTTTTGAAGTACTACTGATTTTATCATGATGAGAGCTCTACAACT	1908
QY	321	HisAspLeuGlnProGlyIleTyrIleTyrProThrValIlePheIleTyrGlyIleProGln	340
DB	1909	CATGATCTACAGCTGGAAAGAAATATCTACTGTGCTGTTCATATATGATGCTCTCAG	1968
QY	340	-----	340
DB	1969	GTCGAGTTGGTGAATATCGTTTAAAGAGTCAAGTATTCCGCTTGAATACCTTAGCC	2028
QY	340	-----	340
DB	2029	TCTCTAGTATATGTTGTTAGTATAGACAAAGGAGATCTGTACAGAGGCTTAA	2088
QY	340	-----	340
DB	2089	TTTGAAGCGCTTTAAATATAAATGGTCAATAGAAATGACGATCAGTGAAGA	2148
QY	340	-----	340
DB	2149	CTCCAAATATCTAGCTTCTGATATGATTTCACTTACTAGATCTGTGGCATCCAGC	2208
QY	340	-----	340
DB	2209	TGATCTATGAGAGATACCTCTCCCTGATGATTAAGCAGAGTCAGATATCTTCAGG	2268
QY	341	ValAlaIleAlaGlyAlaProValIleLeuTyrIlePheTyrAspThrGlyTyrThrGlu	360
DB	2269	GTTGCTATTGCTGGGCGCCAGTCACTGTGATCTTCTATGATACAGATACACGAA	2328
QY	361	ArgTyrMetGlyHisProAspGlnAsnGluGluIleTyrTyrThrLeuGlySerValAlaMet	380
DB	2329	CGTTATATGGGTACCTGACCAAGATGAACAGGCGTATTAAGATCTGGCCATG	2388
QY	381	GlnAlaGlyIleAspPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp	400
DB	2389	CAGCAGAAAGTTCCTCCCTGAAACCAATCGTTTACTGCTCTTACATGGTTCTCGAT	2448
QY	401	GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys	420
DB	2449	GAGAAATGCTCATTTTGCATACACAGTATATCTAGTTTATGAGAGGCTGGGAAG	2508
QY	421	ProTyrAspLeuGlnIleTyrProGlnArgHisSerIleArgValIleProGluSerGly	440
DB	2509	CCATATGATTTTACAGATCTATCTCAAGAGACACAGCATTAAGATCTTGAATGGGA	2568
QY	441	GluHisTyrGluLeuHisIleLeuHisIleTyrLeuGlnGluAsnLeuGlySerArgIleAla	460
DB	2569	GAACATTATGAACATGATCTTTTGGACCTACCTTCAAGAAACCTTGATCAGCTATTGCT	2628
QY	461	AlaLeuLysValIle 465	
DB	2629	GCTCTAAAGTGATTA 2643	

## RESULT 7

ABX12255 standard; cDNA; 2649 BP.

ABX12255;

19-MAY-2003 (first entry)

CDNA encoding human serine protease H1PHM46.

Human; ss; gene; H1PHM46; serine protease; gene therapy; osteoarthritis;  
 serine protease activity modulation; dipeptidyl peptidase activity;  
 musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;  
 Alzheimer's disease; parsynuclear palsy; Huntington's disease;  
 amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;



KW irritable bowel syndrome; type 1 diabetes; faecal incontinence; anaemia  
KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;  
KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;  
KW multiple sclerosis; chromosome 15q21-q22.

OS Homo sapiens.

	Key	Location/Qualifiers
FH	CDS	1..2649
FT		

PN GB2374869-A.

PD 30-OCT-2002.

PF 22-JAN-2002; 2002GB-0001404.

PR 23-JAN-2001; 2001GB-0001760.

PA (GLAX ) GLAXO GROUP LTD.

PI Edbrooke MR, Lewis AP;

DR WPI: 2003-150703/15.

DR P-P;DB; ABU07720

PT Identifying modulators of serine protease activity useful for treating  
PT muscular/skeletal diseases, by contacting cell expressing a novel serine  
PT protease polypeptide with a compound and monitoring serine protease  
PT activity -

PS Claim 12; Page 22-26; 38pp; English.

The invention relates to a method of identifying a substance that modulates serine protease activity, comprising contacting a cell, such as a neuronal cell, lung cell, intestinal cell or a cell infected with a virus, expressing a serine protease polypeptide (HHPNM46), or its variant having dipeptidyl peptidase activity, or a serine protease isolated from the cell with a test substance and monitoring for serine protease activity. The method is useful for identifying a substance that modulates serine protease activity. A modulator of the serine protease is useful in the manufacture of a medicament for treatment or prophylaxis of a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus infection, Alzheimer's disease, paraspontaneous palsy, myotonic dystrophy, Huntington's disease or amyotrophic lateral sclerosis. Additional diseases that may be treated using modulators of the serine protease include malabsorption syndromes, irritable bowel syndrome, lung disease, type II diabetes, faecal incontinence, haemorrhoids, proctitis, rectal polyps, small bowel tumours, colorectal tumours, anaemia, dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple sclerosis. The present sequence represents cDNA encoding the human serine protease HHPNM46 which is located on chromosome 19q21-q22.

SQ Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;

### Alignment Scores:

Pred. No.:	5,82e-248	Length:	2649
Score:	2422.00	Matches:	465
Percent Similarity:	82.30%	Conservative:	0
Best Local Similarity:	92.30%	Mismatches:	0
Query Match:	97.58%	Indels:	100
DB:	25	Gaps:	1

US-10-070-464-5 (1-465) X ABX12255 (1-2649)

QY 1 ThrGlyTTrAlaAspProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
Db 952 ACAGGTACGCGAAATCTTAAAGTCACTTTTAAAGATGTAGAAATATATATTGATGCTCAA 1011  
QY 21 GLVArgIleIleLeaspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40

Db	1012	GGAAGAGCATGATAGTCATAGATTAAGGAACATACTTTGAGATTCATTGGA	107
QY	41	GIYVALGIUTYRIEALARGALAGIYTPThrProGluGIYVSYFALATrSerIle	60
Db	1072	GGAGTTGAATATATATGGCCAGAGCTGAGTGAAGTCTCTGAGGAGAAATATGCTTGTCATTC	113
QY	61	LeuLeuAPArGserGIInThrArgGluInIleValLeuIleSerProGluLeuPheIle	80
Db	1132	CTACTAGATCCCTCCCAAGCTCGCTGGAGATAGTGGTGAATCTCACTGAATATATTTATAC	119
QY	81	ProValGIuAPArSerValMetGIuArgGIInArgLeuIleGluSerValProAspSerVal	100
Db	1192	CCAGTGAAGATGATGATTTATGAGAAAGCAGAGACTCATTTAGTCACTGCTGATCTTGIG	125
QY	101	ThrProLeuIleIleIerYrGluGIuThrThraSPIetpRIleAsnIleHisAspIlePhe	120
Db	1252	AAGCCATAATATCTATGAGAAACAACAGACATCTGGATTAATATCCATGACATCTTT	131
QY	121	HisValPhePProGluInSerHisGluGluGIuIleGluPheIlePheAlaSerGIuCyIysV	140
Db	1312	CATGTTTTTCCCAAGTCAAGAGAGGAAATTGATTATTTTCCCTGGAATGCAAA	137
QY	141	ThrgIYpHeArGHIleuYrLyIleYrIleThrserIleLeuIleuVbGluSerIlystYrIysArg	160
Db	1372	ACAGTTTCCGTCATTATACAAAATTACATCTATTTTAAAGGAAAGCAAAATATAACGA	143
QY	161	SerSerGIYGIYLeuProAlaProSerAspPheIysCysProIleIeGIuGIuIleAla	180
Db	1432	TCCAGTGGTGGCGCTGCCGTCCAAGTATTTCAAAGTCTCTATCAAAAGAGAGATACGA	149
QY	181	IleThrserGIYGIuTPRGIuValIleuGIYArGHIsgIYSerAsnIleGlnValAspGIu	200
Db	1492	ATTACCAAGTGTGATGGAGAGTTCTTGCCGCGCATGATCTAAATATCCAAGTTGATGAA	155
QY	201	ValArgArgLeuValIYrPheGluGIYThrIYsAspSerProLeuGIuHisIleuYr	220
Db	1552	GTCAGAAAGCTGTGATATTTGAAAGGCACCAAGACTCCCTTTAGACATCACCTGTAC	161
QY	221	ValValSerYrValAsnProGIYGIuValThraArgLeuThraAspArgGIYTYSerHis	240
Db	1612	GTACTCACTTACGTAATCTCGAGAGGTGACAAAGCTGACTGACCGTGGCTACTACAT	167
QY	241	SerCysCysIleSerGlnHisCysAspPhePheIleSerIYrYSerAsnGlnIYsAsn	260
Db	1672	TCTTGCTGCATCAGTCAGACTGTGACTCTTTATTAAGTATAGTATACAGAAAGAT	173
QY	261	ProHisCysValSerLeuYrIYleuLeuSerSerProGIuAPArProThrCyIysThr	280
Db	1732	CCACAAGTGTGCTCCCTTTACACCTATCAAGTCCCTGAAGATGACCCAACTTGCAAAACA	179
QY	281	LYeGIuPheTPAlaThrIleLeuAspSerIAGIYProLeuProAspTYrThrProPro	300
Db	1792	AAGGAATTTTGGGCACCACTTTTGATTCACACAGTCTCTTCCGCTGAATATCTCTCA	185
QY	301	GluIlePheSerPheGluSerThrThnGIYpHeThrLeuYrGIYMeLeuYrIYsPro	320
Db	1852	GAAATTTTCTCTTTGAAAGTACTACTGGAATTTACATTGATGGAGTCTTACAAAGCT	191
QY	321	HisAspLeuGIuProGIYLYeLYeYrYrProThrValLeuPheIleYrGIYGIYProGlu	340
Db	1912	CATGATCTACAGCTGGAAGAAATATCTACTGTGCTTCATATATAGTGTGCTCAG	197
QY	340	-----	340
Db	1972	GTCGAGTTGCGTAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTAGCC	203
QY	340	-----	340
Db	2032	TCTCTAGGTTATGCGTTGTAGTATAGACAACAGGGAGATCTCTCAACGAGGCTTAAA	209
QY	340	-----	340
Db	2092	TTTGAAAGCGCTTTAAATATATAAATGGGTCAATAGAAATTGACGATCAGGTGAAGA	215

```

QY 340 ----- 340
Db 2152 CTCCAATATAGCTTTCGATATGATTTGACTAGATCGTGGGATCCACGGC 2211
QY 340 ----- 340
Db 2212 TGGTCCATAGAGGATACCTCTCCCTGATGGCATTATACAGAGTCAGATATCTTCAGG 2271
QY 341 ValAlaIleAlaGlyAlaProValThirLeuTrpIlePheTyrAspThrGlyTyrThrGlu 360
Db 2272 GTTGCTATTGCTGGGGCCCAAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAA 2331
QY 361 ArGlyTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMet 380
Db 2332 CGTATATATGGGTCACCTCGACCCAGATACACAGGGCTTATCTTAGAGATCTGTGGCCCATG 2391
QY 381 GlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400
Db 2392 CAAAGCAAAAAGTTCCCTCTCGAACCAATCGTTTACTGCTCTTACATGATGTTTCTTGAT 2451
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420
Db 2452 GAGATATGCCATTTTGCACATACCGATATATTACTGAGTTTATAGTAGGGGCTGGAAAG 2511
QY 421 ProTyrAspLeuGlnIleTyrProGlnGlnLysSerIleArgValProGlnSerGly 440
Db 2512 CCATATGATTTTACAGATCTATCTCAGAGACACACATTAAGATTCCTGAATCTGGA 2571
QY 441 GluHisTyrGlnLeuHisLeuLeuHisTyrLeuGlnGlnAsnLeuGlySerArgIleAla 460
Db 2572 GAACATTTATGAACGCACTCTTTGCACTACCTTCAAGAAAACCTTGGATCAGATTTGCT 2631
QY 461 AlaLeuLysValIle 465
Db 2632 GCTCTAAAGTGATA 2646

RESULT 8
ABK83322
ID ABK83322 standard; cDNA; 2671 BP.
XX
AC ABK83322;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPPIV related serine protease DPP-1.
XX
KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200231134-A2.
XX
XX 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US31874.
XX
XX 12-OCT-2000; 2000US-240117P.
XX
XX (FERR ) FERRING BV.
XX
XX Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
XX WPI: 2002-444178/47.
XX
XX P-FSDB; ABG61591.
XX
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding

```

```

PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
PS Claim 1; Page 53-54; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPPP proteins.
XX
SQ Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other;

```

## Alignment Scores:

```

Pred. No.: 5,896-248 Length: 2671
Score: 2422.00 Matches: 465
Percent Similarity: 82.30% Conservative: 0
Best Local Similarity: 82.30% Mismatches: 0
Query Match: 97.58% Indels: 100
DB: 24 Gaps: 1

```

US-10-070-464-5 (1-465) x ABK83322 (1-2671)

```

QY 1 ThrGlyThraIaenProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
Db 959 ACAAGTACAGCAAAATCTTAAGTACCTTTTAAGTACAGAAATATGATGATGCTGAA 1018
QY 21 GlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGluIleLeuPheGlu 40
Db 1019 GGAAGGATCATAGATGTCATATGAAAGAACTAATTCACCTTTGAGATTCATTTGAA 1078
QY 41 GlyValGlnTyrIleAlaArgAlaGlyTyrThrProGlnGlnLysTyrAlaTrpSerIle 60
Db 1079 GGAATTAATATATTCGACAGCTGATGAGCTCTCAGGAAATATGCTGGTCCATTC 1138
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
Db 1139 CTACTAGATGCGCTCCAGAGCTCGCTGCAGATAGTTGATCTCAGCGAATTTATATC 1198
QY 81 ProValGlnAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
Db 1199 CCACTAGAAAGATGATGTTATGGAAGGACAGAGACTCATTTGAGTCAGGCGCTGATTCGTG 1258
QY 101 ThrProLeuIleIleTyrGlnGlnGluThrThrAspIleTrpIleAsnIleHisAspIlePhe 120
Db 1259 ACGCCACTAATTAATCTATGAGAAACAACAGACATCTGATTAATATCCATGACATCTTT 1318
QY 121 HisValPheProGlnSerHisGlnGlnGlnIleGluPheIlePheAlaSerGluCysLys 140
Db 1319 CATGTTTTTCCCAAAAGTCAGAAAGAGAAATTGATTTATTTTGGCTCTGAAATGCAAA 1378
QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGlnSerLysTyrLysArg 160
Db 1379 ACAGGTTTCGTCATTTATACAAATTAACATCTATTTTAAAGGAACCAATATAAACA 1438
QY 161 SerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGlnGluIleAla 180
Db 1439 TCCAGTGTGGGCTGCTCTCCAGATGATTTTCAAGTCTCTATCAAGAGAGATGACA 1498
QY 181 IleThrSerGlyGlnTyrGlnValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200
Db 1499 ATTACAGTGTGATGGAAGATTTCTTGCCGCGATGATCTAATATCAAGTGAATGAA 1558

```

```

QY 201 ValArgArgLeuValIYrPheGluGlyThrLysAspSerProLeuGluHISLeuTYr 220
DB 1559 GTGGAAGGCTGGATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGGAC 1618
QY 221 ValValSerTYrValAaenProGlyGluValThrArgLeuThrAaPAGLYrTYrSerHis 240
DB 1619 GTAGTCGATGACGTAATCCCTGAGAGGTGCAAGGCTGATGACCTGGCTACTGCAT 1678
QY 241 SerCysAlleserGlnHisCysAspPhePheIleSerLysTYrSerAaenGlnLysAaen 260
DB 1679 TCTTGCTGCATCACTGACACTGACTCTTTTAAAGTAAGTATAGTAAACCAAGAAAT 1738
QY 261 ProHisCysValSerLeuTYrLysLeuSerSerProGluAaPProThrCysLysTYr 280
DB 1739 CCACACTGTGTGCTCTTTCAAGCTATCAAGTCTTCAAGATGACCACTGCAAAACA 1798
QY 281 LysGluThrPheThrIleLeuAspSerAlaGlyProLeuProAaPYYrTYrProPro 300
DB 1799 AAGGAATTTGGGCGACCATTTGGATTCAGCAGGCTCTCTCTGACTATACTCCCA 1858
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysPro 320
DB 1859 GAAATTTTCTCTTTGAAAGTACTGACTGATTTTCACTGATGAGGATGCTTACAAAGCT 1918
QY 321 HisAspLeuGlnProGlyLysLysTYrProThrValLeuPheIleTYrGlyGlyProGln 340
DB 1919 CATGATCTACAGCTGGAAAGAAATATCTGCTGCTCATATATGGTGGTCTCAG 1978
QY 340 ----- 340
DB 1979 GTGCACTGTGTGAATATCATTTTAAAGAGTCAAGTATTCGGCTTAAATACCTTACC 2038
QY 340 ----- 340
DB 2039 TCTCTAGTATATGTTGTTAGTAGATAGACAAAGGGATCTCTACCAAGGGCTTAAA 2098
QY 340 ----- 340
DB 2099 TTTGAAGGCGCTTTAAATATAAATGGTCAATAGAAATGACGATCAGTGAAGAA 2158
QY 340 ----- 340
DB 2159 CTCCAATATCTAGCTTGCATATGATTTTCACTTGAATGATGCTGTGGCATTCACGCG 2218
QY 340 ----- 340
DB 2219 TGGTCTATGAGAGATACCTCTCCCTGATGAGCATTAATGACAGAGTACGATATCTTACG 2278
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTYrAspThrGlyTYrThrGlu 360
DB 2279 GTTGCTATGCTGGGGGCCAGTCACTGTGATCTTCTATGATACAGGATACAGGAA 2338
QY 361 ArgTYrMetGlyHisPheAspArgIAsnGluGlyTYrTYrLeuGlySerValAlaMet 380
DB 2339 CGTTATATGGGTACCTCTGACCAAGATGAACAGGGCTATTAAGATCTGTGGCCATG 2398
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuHISGlyPheLeuAsp 400
DB 2399 CAAGCAGAAAAGTTCCCTCTGAACCAAATCGTTTACTGCTTCACTGATGTTCTGGAT 2458
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420
DB 2459 GAGAAATGTCATTTTGACATACACAGATATTAAGTATTTTAAAGAGGCTGGAAG 2518
QY 421 ProTYrAspLeuGlnIleTYrProGlnGluArgHisSerIleArgValProGluSerGly 440
DB 2519 CCAATATGATTTTAAAGATCTATCTCAGGAAGACACAGCAATTAAGAGTCTTGAATCGGGA 2578
QY 441 GluHisTYrGlyLeuHisLeuLeuHisTYrLeuGlnGluAsnLeuGlySerArgIleAla 460
DB 2579 GAACATTAATGAACCTGATCTTTTGGCACTACCTTCAAGAAAACCTTGGATCAGCTATTTGCT 2638
QY 461 AlaLeuLysValIle 465

```

```

DB 2639 GCTCTAAAGTGATA 2653
RESULT 9
ABK12892
ID ABK12892 standard; cDNA; 3106 BP.
AC ABK12892;
XX 09-APR-2002 (first entry)
XX Human protease PRTS-9 cDNA sequence.
DE
XX Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KM cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KM inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KM cell proliferative disorder; developmental disorder; epilepsy;
KM Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KM reproductive disorder; endometriosis; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 203..2851
XX FT 0 /*tag= a
XX FT /*product= "Human protease PRTS-9"
XX PN MO200198468-A2.
XX
XX 27-DEC-2001.
XX
XX 13-JUN-2001; 2001MO-US19178.
XX
XX 16-JUN-2000; 2000US-212336P.
XX 22-JUN-2000; 2000US-213955P.
XX 29-JUN-2000; 2000US-215396P.
XX 07-JUL-2000; 2000US-216821P.
XX 14-JUL-2000; 2000US-218946P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM;
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Malia NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Rankumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
XX WPI; 2002-090437/12.
XX P-PSDB; AAU74749.
XX
XX Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
XX Claim 5; Page 166-167; 177pp; English.
XX
XX The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present nucleic acid sequence encodes
CC the human protease PRTS-9 protein of the invention.
XX
XX Sequence 3106 BP; 928 A; 633 C; 704 G; 841 T; 0 other;
SQ

```



```

FT      /*tag= a
XX      /product= "Human DPP8"
PN      MO200119866-A1.
XX      22-MAR-2001.
XX      11-SEP-2000; 2000MO-AU01085.
XX      10-SEP-1999; 99AU-0002762.
PR      18-FEB-2000; 2000AU-0005709.
XX      (UNSY ) UNIV SYDNEY.
XX      Abbott CA, Gorell MD;
PI      MPI; 2001-281520/29.
XX      P-PSDB; AAA87187.
XX      New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT      substrates, identifying inhibitors of DPP8 catalytic activity which
PT      have therapeutic uses, and for detecting activated T cells
XX      Claim 16; Fig 2; 78pp; English.
XX      This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8
CC      has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and
CC      H-Arg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a
CC      dipeptidyl peptidase, because it is capable of hydrolyzing the
CC      peptide bond C-terminal to Pro in each of these compounds. DPP8
CC      is homologous with human DPPIV. DPP8 is useful for cleaving a
CC      substrate, and for detecting an activated T cell which involves
CC      measuring the level of DPP8 gene expression in a T cell. The level
CC      of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC      in the cell. It is also useful for identifying a molecule capable
CC      of inhibiting the cleavage of the substrate by DPP8. Molecules
CC      identified as inhibiting DPP8 catalytic activity may be useful for
CC      treating diabetes, growth hormone deficiency, lowering glucose levels
CC      in non-insulin dependent diabetes mellitus and other disorders
CC      involving glucose intolerance, enhancing mucosal regeneration and
CC      as immunosuppressants.
SQ      Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
Score: 7,45e-248 Length: 3120
Percent Similarity: 2422.00 Matches: 465
Best Local Similarity: 82.308 Conservative: 0
Query Match: 97.584 Mismatches: 0
DB: 22 Indels: 100 Gaps: 1

US-10-070-464-5 (1-465) x AAC85694 (1-3120)
QY      1 ThrGlyThrAlaAsnProLySValThrPheLySMeLSerGluLeuMetIleAspAlaGlu 20
DB      1165 ACAGGTACGCAAAATCTTAAGTCACTTTTAAGATGTCAGAAATATATGATGTCGTA 1224
QY      21 GlyAglIleIleAspValIleAspLySgluLeuIleGlnProPheGluIleuPheGlu 40
DB      1225 GGAAGGATCATAGATGTCATAGATAGAACTAATTCACCTTTGAGATTCATTTGAA 1284
QY      41 GlyValIGluTyrlleAlaArgAlaGlyTrpThrProGluGlyLySerTyrlaIlePheSer 60
DB      1285 GGAAGGATCATAGATGTCATAGATAGAACTAATTCACCTTTGAGATTCATTTGAA 1284
QY      61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleuIleSerProGluLeuPheIle 80
DB      1345 CTACTAGATGCTCCGACAGCTGCGCTACAGATAGTGTGATCTCACCTGAATTTATTTATC 1404
QY      81 ProValIGluAspAspValMetGluArgGluArgLeuIleGluSerValProAspSerVal 100
DB      1405 CCAAGTAAGATGATGTTATGAGAAAGGACAGACCTCATTTGAGTGCAGTCTGCTGTG 1464

```

```

QY      101 ThrProLeuIleIleTyrgluGluThrThrAspIleTrpIleAsnIleHisAspIlePhe 120
DB      1465 ACGCCACTAATTTATCTATGAGAAACACAGACATCTGATTAATATTCATGACATCTTT 1524
QY      121 HisValPheProGlnSerHisIleGluGluIleGluPheIlePheIleAspGluCysLys 140
DB      1525 CATGTTTTCCCAAGAGTCACGAGAGGAAATTGAGTTTATTTTCCCTCTGATGAGAA 1584
QY      141 ThrGlyPheArgHisIleuTyrlYlylleThrSerIleLeuLySgluSerLyTyrlYsArg 160
DB      1585 ACAGGTTTCGCTCATTTATACAAATTAATCATCTATTTTAAGGAAAGCAATTAACGA 1644
QY      161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLySgluGluIleAla 180
DB      1645 TCCAGTGTGGGCTGCTGCTCCAGATGATTCAGAGTCTCATCAAGAGGAGATAGCA 1704
QY      181 IleThrSerGlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGluValAspGlu 200
DB      1705 ATTACCAAGTGGTGAATGGGAAGTTCTGGCCGSCATGATCTTAATTCAGATTGATGAA 1764
QY      201 ValArgArgLeuValIlyrPheGluGlyThrLySAspSerProLeuGluHisIleuTyrl 220
DB      1765 GTCAGAGGCTGGTATATTTTGAAGGACCAAAAGCTCCCTTTAGAGCATCAGCTGTAC 1824
QY      221 ValValSerTyrlValAsnProGlyGluValIleThrArgLeuThrAspArgGlyTyrlSerHis 240
DB      1825 GTAGTCACTAGCTAATATCTGAGAGGTGACAGGCTGATCAGCCGTGCTATCATACAT 1884
QY      241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLySerAsnGlnLySAsn 260
DB      1885 TCTGTGATCAGTACGACAGCTGATCTTTATAGTATGATATGTAACCAAGAAAT 1944
QY      261 ProHisCysValSerIleuTyrlYsLeuSerSerProGluAspAspProThrCysLyThr 280
DB      1945 CCACACTGTGTCTCTTTCAAGCATATCAAGCTCAGAGATGACCAACTTCAGAAACA 2004
QY      281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrlThrProPro 300
DB      2005 AAGAAATTTGGCCACCAATTTGATGACAGAGCTCTCTTCGACATATCTCTCCCA 2064
QY      301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrlYMetLeuTyrlYsPro 320
DB      2065 GAAATTTCTCTTTGAAAGTACTGATTTTACTTATGAGATGCTCTCAAGGCT 2124
QY      321 HisAspLeuGlnProGlyLySlySlyTyrlProThrValLeuPheIleTyrlGlyProGln 340
DB      2125 CATGATCTACAGCTCGAAAGAAATATCTACTGTGCTGTTCATATATGATGCTCTCAG 2184
QY      340 ----- 340
DB      2185 GTTCAGTTGGTGAATTAATCGTTTAAGAGATCAAGATATTCGCTTGAATACCTTACC 2244
QY      340 ----- 340
DB      2245 TCTCTAGTTATGTTGTTGTTAGTATGATACACACAGGGATCTGTGCCGAGGCTTAA 2304
QY      340 ----- 340
DB      2305 TTTGAAGGCGCTTTAAATATTAATGAGTCAATATAGAAATTCAGATCAGGTGAAAGA 2364
QY      340 ----- 340
DB      2365 CTCCAATATCTAGCTTCTCGATATGATTTGATCTTAGATCGTGTGGCATTCACGCGC 2424
QY      340 ----- 340
DB      2425 TGTGCTATGAGGATACCTCTCCCTGATGGCATTAAATGACAGAGTCAGATATCTTCAGG 2484
QY      341 ValAlaIleAlaGlyValaProValThrLeuTrpIlePheTyrlAspThrGlyTyrlThrGlu 360
DB      2485 GTTGTCTATGCTGGGCCCCAGTCACTGTGTGATCTTCTATGATACAGGATACAGGAA 2544

```

QY 361 ArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 Db 2545 CGTATATGGGTCACTGACCGAATGACAGGGCTATTACTAGGATCTGTGGCATG 2604  
 QY 381 GlnAlaGlnLysPheProSerGluProAsnArgLeuLeuHisGlyPheLeuAsp 400  
 Db 2605 CAAGCAGAAAAGTCCCTCTGAAACCAATGTTTACTGCTTACATGCTTCTGGAT 2664  
 QY 401 GlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 Db 2665 GAGAAATCTCCATTGTGCACATACAGATATATTACTGAGTTTAAAGTGAAGGCTGGAAG 2724  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGluSerGly 440  
 Db 2725 CCATATGATATTACAGATCTATCTCTCAGAGACACAGCATAAAGTTCTCGAATCCGGA 2784  
 QY 441 GlnHisTyrGlnLeuHisIleLeuHisTyrLeuGlnGlnLeuHisGlySerArgIleAla 460  
 Db 2785 GAACATTATGACATGCTCTTTGTCACTCACTCAAGAAAACCTTGAGTACGTAATGCT 2844  
 QY 461 AlaLeuLysValIle 465  
 Db 2845 GCTCTAAAGTGATA 2859

## RESULT 11

AAD38956  
 ID AAD38956 standard; cDNA; 3120 BP.

AC AAD38956;

DT 23-SEP-2002 (first entry)

XX Human dipeptidyl peptidase 8 (DPP8) cDNA.

XX Human dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;

KM autoimmunity; human immuno deficiency virus; HIV infection; cytosolic;

KM graft rejection; antidiabetic; antiinflammatory; immunosuppressive;

KM antiviral; enzyme; gene; ss.

XX Homo sapiens.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC

CC sequence is human DPP8 cDNA.  
 XX  
 SQ Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;  
 Alignment Scores:  
 Pred. NO.: 7.45e-248 Length: 3120  
 Score: 2422.00 Matches: 465  
 Percent Similarity: 82.30% Conservative: 0  
 Best Local Similarity: 82.30% Mismatches: 0  
 Query Match: 97.58% Indels: 100  
 DB: 24 Gaps: 1  
 US-10-070-464-5 (1-465) x AAD38956 (1-3120)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 1165 ACAGGTACAGCAAAATCTTAAGTACCTTTTAAGATGTCAGAAATTAATGATTGATCGGAA 1224  
 QY 21 GlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 1225 GGAAGCATCATAGTGCATAGATTAAGGAATTAATCACTTTTGAGATTCTATTGAA 1284  
 QY 41 GlyValGlnTyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyrAlaTrpSerIle 60  
 Db 1285 GAGTTGAATATATGTCAGAGCTGATGATCTTCAGGAAATATAGCTTGATCCATC 1344  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1345 CTACTGATGCTGCCAGACTCCGCTTACATGATGTTGATCTCAGCTGATTTATTC 1404  
 QY 81 ProValGlnAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1405 CCACTGAAGATGATGTTATGAGAAAGCAGAGACTCACTGAGCAGCTGATTCGTG 1464  
 QY 101 ThrProLeuIleIleTyrGlnGlnThrThrAspIleIleTrpIleAsnIleHisAspIlePhe 120  
 Db 1465 ACGCCTAATATCTATGAAAGAACACAGACTCTGATTAATATTCATGACATCTTT 1524  
 QY 121 HisValPheProGlnSerHisGlnGlnGlnIleGlnPheIlePheAlaSerGlnCysLys 140  
 Db 1525 CATGTTTTTCCCAAGTACAGAAAGGAATGAGTTATTTTGGCTCTGATGCAAA 1584  
 QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 Db 1585 ACAGTTTCCGCTATTAACAATAATCACTTAATTAAGAAACCAATAATMAACA 1644  
 QY 161 SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGlnGluIleAla 180  
 Db 1645 TCCAGTGTGGCTGCTCTCCAAAGTATTCAGATGCTCTATCAAAAGAGAGATAGCA 1704  
 QY 181 IleThrSerGlyGlnTrpGlnValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 Db 1705 ATTACCAAGTGTGATGGAAGTTCTTGGCGGATGATCTAATATCCAAAGTTGAGAA 1764  
 QY 201 ValArgArgLeuValTyrPheGlnGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyr 220  
 Db 1765 GTCAGAAAGCTGTGATATTTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGTAC 1824  
 QY 221 ValValSerTyrValAsnProGlyGlnValThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1825 GTAGTCAGTTACGTAATCTCGAGAGGTGACAAAGGTGACTGACCTGCTACTCAT 1884  
 QY 241 SerCysValIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 Db 1885 TCTTGCTGATCAGTACGACTGTGACTCTTATTAAGTAAATAGTAACTCAGAGAAAT 1944  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGlnAspAspProThrCysLysThr 280  
 Db 1945 CCACACGTGTGTCCCTTCAAGCTATCAAGTCTGAAATATCAACCACTCTCAAAACA 2004  
 QY 281 LysGlnPheThrPheAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 2005 AAGGAATTTTGGGCCACATTTTGGATTCAAGCAGGTCTCTTCTGACTATATCTCTCCA 2064

```

QY 301 Glu11ePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320
DB 2065 GAATTTTCTCTTGGAAAGTACTAGTATTCATTGATGAGGATCTTACAAAGCTT 2144
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340
DB 2125 CATGATCTACAGCTGGAAAGAAATATCTCTAGTGTGCTTATATATGATGATGCTCTAG 2184
QY 340 ----- 340
DB 2185 GTGCAATTGGTGAATTAATCGGTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTAGCC 2244
QY 340 ----- 340
DB 2245 TCTCTAGTATTATGTTGTTAGTATGATACACAGGGATCTCTACAGGGCTTAA 2304
QY 340 ----- 340
DB 2305 TTGGAAGCGCTTTAATATATAATGGGTCAAATAGAAATTGACGATCAGGTGAAGGA 2364
QY 340 ----- 340
DB 2365 CTCCAATATATCTAGCTTCTCGATATGATTTCACTTACCTTACGATCGTGGGCATCCAGGC 2424
QY 340 ----- 340
DB 2425 TGGTCTTATGAGAGTATCTCTCCCTGATGSCATTAAATGACAGAGTCAGATATCTTACAGG 2484
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360
DB 2485 GTTGCTATTTCTGGGGCCCGACGTCACCTGTGAGTCTTTATGATATACAGATACACAGAA 2544
QY 361 ArgTyrMetGlyHisAspAspGlnAsnGlnGlyTyrTyrLeuGlySerValAlaMet 380
DB 2545 CGTTATATGGGTACCTGACCAAGATGAACAGGGCTATTAATGAGATCTGTGGCCATG 2604
QY 381 GlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400
DB 2605 CAAGCAGAAAGTTCCCTCTGAAACCAATTCGTTACTGCTCTTACATGGTTCTTGAT 2664
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLys 420
DB 2665 GAGAAATGTCATTTTGCACATACACAGATATTAAGTATTTTGAAGGGCTGGAAG 2724
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlnSerGly 440
DB 2725 CCATATGATTTTACAGATCTATCTCAGAGAGACACAGCATAGAGTTCTGAAATCGGGA 2784
QY 441 GluHisTyrGlyLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460
DB 2785 GAACTATTATGACCTGCACTCTTTTGCACATCTTCAAGAAACCTTGGATCACGATTTGCT 2844
QY 461 AlaLeuLysValIle 465
DB 2845 GCTCTAAAGTGATA 2859

```

RESULT 12  
AAH99934  
ID AAH99934 standard, cDNA, 3143 BP.

AC AAH99934;

DT 12-APR-2002 (first entry)

DE cDNA encoding 21953 human prollyl oligopeptidase.

XX 21953 prollyl oligopeptidase; human; prollyl; endopeptidase;  
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;  
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;  
KW antidiabetic; antiarthritic; antiatherosclerotic; antiinflammatory;  
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;  
KW Grave's disease; neuronal disorder; demyelinating disease; ss.

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 0 229..2877
FT /tag= a
FT /product= "21953 prollyl oligopeptidase"
FT /note= "This region is specifically claimed in
claim 2"
XX MO200179473-A2.
XX 25-OCT-2001.
XX 11-APR-2001; 2001WO-US40483.
XX 18-APR-2000; 2000US-197508P.
XX (MILL-) MILLENNIUM PHARM INC.
XX PA Meyers RA, Williamson M;
XX WIPI; 2002-034353/04.
XX P-PSDB; AAG78415.
XX New polypeptides 21953, member of human prollyl oligopeptidase family,
XX useful as diagnostic targets and therapeutic agents for controlling
XX cancer, lymphoma and leukemia
XX Claim 7; Page 100-102; 121pp; English.
XX CC This invention relates to an isolated 21953 human prollyl
XX CC oligopeptidase. Which is cytostatic, antidiabetic, antiarthritic,
XX CC neoprotective, antihypoid, dermatological, antipsoriatic,
XX CC antiatherosclerotic, ophthalmological, antiinflammatory, neurotropic,
XX CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
XX CC antianginal, cardiac, antiatherosclerotic, anorectic and
XX CC metabolic in its action. Uses include gene therapy, expression or
XX CC activity of 21953 protein modulator, it is useful for identifying a
XX CC compound which binds to it and can be used in preventing, treating
XX CC or detecting a cellular proliferative or differentiative disorder.
XX CC The 21953 molecules can act as novel diagnostic targets and therapeutic
XX CC agents for controlling disorders associated with the aberrant activity
XX CC or degradation of peptide hormones e.g., disorders associated with cell
XX CC differentiation and proliferation such as cancer, immune function,
XX CC reproductive, neurological and cardiovascular function. The 21953
XX CC molecules are thus useful for treating and preventing cellular
XX CC proliferative and differentiative disorders, haematopoietic neoplastic
XX CC disorders, immune disorders such as autoimmune diseases, diabetes
XX CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
XX CC neuronal disorders, demyelinating diseases, vascular disorders and
XX CC metabolism or pain disorders. This sequence represents the cDNA
XX CC encoding sequence of 21953 human prollyl oligopeptidase.
XX SO Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;

Alignment Scores:
Pred. No.: 7,54e-248 Length: 3143
Score: 2422.00 Matches: 465
Percent Similarity: 82.30% Conservative: 0
Best Local Similarity: 82.30% Mismatches: 0
Query Match: 97.58% Indels: 100
DB: 24 Gaps: 1

US-10-070-464-5 (1-465) x AAH99934 (1-3143)
QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGlnIleMetIleAspAlaGlu 20
DB 1180 ACGAGTACGACCAATCTTAAAGTCACTTTTAAAGATTCGAAATTAATGATGATGCTGAA 1239
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40
DB 1240 GGAAGGATCATGATCTCATAGTAAAGAACTAATTCACACCTTTTGAGATTCTTGA 1299

```

QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaATPSeTle 60  
 Db 1300 GGAGTGAATAATATTCACAGCGTGAATGACCTCCAGGAAATATGCTGCTGCATC 1359  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1360 CTACTAGATGCTCCAGACTCGCTCGAGTAGTGTGATCTTCACCTGAAATTAATTTATC 1419  
 QY 81 ProValGluAspArgValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1420 CCACTAGACATGATGTTATGAAAAGCAGAGACTCATTTAGTCAAGTCCGCTGATTCGTG 1479  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 Db 1480 ACCCCACTAATTATCTATGAAAGAAACAACAGACTCGAGTAAATATCCATGACATCTTT 1539  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleAspSerGluCysLys 140  
 Db 1540 CATGTTTTTCCCAAGTCCAGAAAGAGAAATGAGTTATTTTTTGCCTCTGAATGCATA 1599  
 QY 141 ThrGlyPheArgHisIleuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 Db 1600 ACAGTTTCCGTCATTTATACAAAATTACATCTATTTTAAAGAAACCAATATATAACGA 1659  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 Db 1660 TCCAGTGGTGGCTGCTGCTCCAAAGATTTCAAGTCTTCAAAAGAGAGATGACA 1719  
 QY 181 IleThrSerGlyGluTyrGluValLeuGlnYArgHisGlySerAsnIleGlnValAspGlu 200  
 Db 1720 ATTACCACTGATGATGGAAAGTTCTTGCCGCGATGATCTAATATCCAAAGTTGATGAA 1779  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleuTyr 220  
 Db 1780 GTCAGAAAGCTGGATATTTTGAAGCACAAGCAAGACTCCCTTAAAGCATACCTGTAC 1839  
 QY 221 ValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1840 GTAGTCAGTACGAAATCCGAGAGAGTACAAAGGCTGACTGACGCTGCTACTCAT 1899  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 Db 1900 TCTTGCTGATCATGCTGACACTGACTCTTTAATAAGTATATGATACCAAGAAAT 1959  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspArgProThrCysLysThr 280  
 Db 1960 CCACTAGTGTGCTCTTTACAAAGCTATCAAGTCTGAAAGATGCCAACCTGCAAAACA 2019  
 QY 281 LysGluPheThrPheIleThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 2020 AAGGAATTTGGGCGACCATTTGGATTGACAGAGTCTCTCTGACTATACTCTCCCA 2079  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 Db 2080 GAAATTTTCTTTTGAAGTACTACTGATTTACATGATGGAGCTCTACAAAGCT 2139  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 Db 2140 CATGATCTACAGCTGGAAGAAATATCTATCTGTCTCATATATGTTGTTGCTCCAG 2199  
 QY 340 ----- 340  
 Db 2200 GTGCACTGTGTAATAATCGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTAGCC 2259  
 QY 340 ----- 340  
 Db 2260 TCTCTAGTTATGTGTTAGTATAGTACAAACAGGGGATCTGTCAACGAGGGCTTAA 2319  
 QY 340 ----- 340  
 Db 2320 TTTGAAGCGCCTTTAAATATTAATGGTCAATATAGAAATGACATCAGTGCAGTGAAGA 2379

QY 340 ----- 340  
 Db 2380 CTCGAATATCTAGCTTCTGATATGATATTCATTGACTAGATCGTGGCATCCAGCGC 2439  
 QY 340 ----- 340  
 Db 2440 TGGTCTATGAGAGATACCTCTCCCTGATGACATTAATGACAGTCCAGATCTTCAGG 2499  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 Db 2500 GTTGCTATTGCTGGGCGCCAGTCACTCTGTGATCTTTATGATACAGGATACAGGAA 2559  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 Db 2560 CGTTATATGGTACCCCTGACCAAGATGAACAGAGCTATTAATTAGATCTGGCCATG 2619  
 QY 381 GluAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 Db 2620 GAGGATGTCATTTTGCACATACCAAGTATATCTAGTGTCTTTAGTGAAGGCTGGAAG 2679  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 Db 2680 GAGATGTCATTTTGCACATACCAAGTATATCTAGTGTCTTTAGTGAAGGCTGGAAG 2739  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 Db 2740 CCATATGATTTTACAGATCTATCTCAGAGAGACACAGCATTAAGTTCCTGAAATCGGA 2799  
 QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLysGlnGluAsnLeuGlySerArgIleAla 460  
 Db 2800 GAAATATATGACATGATCTTTTGCATCTACCTTCAAGAAAACCTTGATCAGTATTTGCT 2859  
 QY 461 AlaLeuLysValIle 465  
 Db 2860 GCTCTAAAAGTGATA 2874  
 RESULT 13  
 ABR3331  
 ID ABR3331 standard; cDNA; 4676 BP.  
 AC ABR3331;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 XX cDNA encoding human DPR-1 splice variant #7.  
 DE  
 XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPR;  
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KM dyskinetia; reproductive disorder; inflammatory disorder;  
 KM metabolic disorder; gene; ss.  
 OS Homo sapiens.  
 XX  
 XX MO200231134-A2.  
 XX  
 XX 18-APR-2002.  
 PD  
 XX  
 PF 12-OCT-2001; 2001MO-US31874.  
 XX  
 XX 12-OCT-2000; 2000US-240117P.  
 PR  
 XX (FERR) FERRING BV.  
 PA  
 XX QJ S, Akinsanya KO, Riviere PJ, Junien J;  
 XX WPI; 2002-444178/47.  
 DR P-PSDB; ABG61600.  
 XX  
 XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and



PT viral infections, cancers, allergies, neurological disorders, or pain  
 XX -  
 PS Disclosure; Page 72-73; 113pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human serine  
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
 CC proteins (DPP). The dipeptidyl peptidase IV-related proteins (DPP)  
 CC and nucleic acids encoding them are useful for treating infections  
 CC such as fungal, bacterial, protozoan and viral infections, particularly  
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
 CC hypertension, urinary retention, osteoporosis, angina pectoris,  
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
 CC schizophrenia), and dyskinesias. These may also be used in discovering  
 CC therapeutic agents for the treatment of reproductive, inflammatory and  
 CC metabolic disorders. ABK83322-ABK83343 encode human DPP proteins.  
 XX

Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;

Alignment Scores:  
 Pred. No.: 3,256-247 Length: 4676  
 Score: 2418.50 Matches: 463  
 Percent Similarity: 90.10% Conservative: 1  
 Best Local Similarity: 89.90% Mismatches: 1  
 Query Match: 97.44% Indels: 51  
 DB: 24 Gaps: 1

US-10-070-464-5 (1-465) X ABK83331 (1-4676)

QY 1 ThrGlyThrAlaenProLysValThrPhelysmetserGluilemetleaspAlaGlu 20  
 DB 1165 ACAGGTACAGCAATTCCTAAAGTCATTTTAAGATGTCAGAAATAAGATGATGCGAA 1224  
 QY 21 GlyArgileileaspValileaspLysGluLeuileglnProPheGluileLeuPheGlu 40  
 DB 1225 GGAAGATCATAGATGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1284  
 QY 41 GlyValGluTyrileAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerile 60  
 DB 1285 GGAAGTAAATATATATGTCAGAGCTGATGATGATGATGATGATGATGATGATGATG 1344  
 QY 61 LeuLeuAspArgserGlnThrArgleuGlnileValleuileserProGluLeuPheile 80  
 DB 1345 CTACTAAGTCGCTCCAGACTCGCTCAGATAGTGTGATCTCAGCTGATATTTATC 1404  
 QY 81 ProValGluAspAspValmetGluArgGlnArgleuileglnuserValProaspSerVal 100  
 DB 1405 CCAAGTAGAAGATGATGTTATGAAAGGAGAGACTCATTTGATGATGATGATGATGATG 1464  
 QY 101 ThrProleuileileTyrGluGluThrThrAspilleTyrPheleuileHleAspillePhe 120  
 DB 1465 ACCGCACTAATTAATCTATGAAAGAAACAACACATCTCGATTAATATCCATGACATCTT 1524  
 QY 121 HisValPheProGlnSerHisGluGluGlnileGluPheilePheAlaSerGluCysVal 140  
 DB 1525 CATGTTTTTCCCAAGTCAGAAAGGAAATTAAGTTATTTTGGCTCTGAATGCAAA 1584  
 QY 141 ThrGlyPheArgHisleuTyrLysileThrSerilleuLysGluSerLysTyrLysArg 160  
 DB 1585 ACAAGTTCCGTCATTATACAAATTAATCATCTTTTAAAGAAACAAATATAACGA 1644  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProilleLysGluGluileAla 180  
 DB 1645 TCCAGTGTGTGGCTGCTGCCAAGATTTCAAGGTCTTATCAAAAGAGGATGCA 1704  
 QY 181 IleThrSerGlyGluTyrPheGluValleuGlyArgHisGlySerAsnileGlnValAspGlu 200  
 DB 1705 ATTAACGATGTGATGGGAAGTTCTTGGCCGCGATGATCTAATATCCAAAGTTGATGAA 1764  
 QY 201 ValArgArgleuValTyrPheGluGlyThrLysAspSerProleuGlnHisleuTyr 220

DB 1765 GTCAAGAGCTGATATTTTGAAGGACCAAAAGCTCCCTTTAGACATCACTGATAC 1824  
 QY 221 ValValSerTyrValaenProGlyGluValThrArgleuThrAspArgGlyTyrSerHis 240  
 DB 1825 GTACTGATTAATGTAATTCCTGGAGAGTGCAGAAAGCTGACCTGCGCTACTACAT 1884  
 QY 241 SerGlyCysileserGlnHisCysAspPhePheileSerLysTyrSerAsnileLysAsn 260  
 DB 1885 TCTTGTCGATCACTGACGACTGTGACTCTTTTAAGTAAGTAATGATACCGAAGAT 1944  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 1945 CCAACAGTGTGTCTCTTTCAACACTATCAAGTCTGAAGATGACCAACTTGCAGAAACA 2004  
 QY 281 LysGluPheTyrAlaThrilleLeuAspSerAlaGlyProleuProAspTyrThrProPro 300  
 DB 2005 AAGCAATTTTGGCGACCATTTGATTCAGACAGTCTCTTCTGCTATATCTCTCCA 2064  
 QY 301 GluilePheSerPheGlnuserThrThrGlyPheThrLeuTyrGlymetLeuTyrLysPro 320  
 DB 2065 GAAATTTCTCTTTTGAAGTACTGATGATTTATCATTTATGGAATGCTCTACAAAGCT 2124  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValleuPheileTyrGlyLysPro--- 339  
 DB 2125 CATGATCTACAGCTGGAAGAAATATCTTATGCTGCTTCAATATGCTGCTG- GCT 2183  
 QY 339 ----- 339  
 DB 2184 CAATAGAAATTCAGATGATGAGAGCACTCCATATATGATGATGATTC 2243  
 QY 339 ----- 339  
 DB 2244 ATGACTTAAGATGCTGTGGCATTCACGCTGCTCTATGAGAGATACCTCTCCGATG 2303  
 QY 340 -----GlnValAlaileAlaGlyAlaProValThrleu 350  
 DB 2304 GCATTAATGACAGATCATATCTTCAAGGCTTCTATGCTGGGCTCCAGTCACTCTG 2363  
 QY 351 TyrillePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlu 370  
 DB 2364 TGGATCTTCTATGATACAGATACACGAAAGTATATGGGTGACCTGACAGAAATGAA 2423  
 QY 371 GlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProAspSerGluProAsn 390  
 DB 2424 CAGGCTAATTAATTAAGATGTGTGGCCATCAAGCAAAAGTCCCTCTGAACCAAT 2483  
 QY 391 ArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerile 410  
 DB 2484 CGTTTACTGCTCTTACATGTTCTCTGATGAGATGTCCATTTTGACATACCAATATA 2543  
 QY 411 LeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnileTyrProGlnGlu 430  
 DB 2544 TTAAGTATTTTATGAGAGGCTGGAAGCCATATGATTTACAGATCTATCTCAGAG 2603  
 QY 431 ArgHisSerilleArgValProGlnSerGlyGlnHisTyrGluLeuHisleuHisTyr 450  
 DB 2604 AGACACAGCATTAAGATCTCTGAATCGGAGAAATATATACATCTTTTGGACATAC 2663  
 QY 451 LeuGlnGluAsnLeuGlySerArgileAlaAlaLeuLysValile 465  
 DB 2664 CTTCAAGAAACCTTGGATCACTATTTGCTCTAATAAGTATTA 2708  
 RESULT 14  
 ABK83327 standard; cDNA; 4829 BP.  
 ID ABK83327  
 XX  
 AC ABK83327;  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DPP-1 splice variant #3.  
 XX

KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPP; DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain; diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; heart failure; hypertension; urinary retention; osteoporosis; cancer; ulcer; allergy; cancer; psychotic disorder; neurological disorder; dyskinesia; reproductive disorder; inflammatory disorder; metabolic disorder; gene; ss.

XX Homo sapiens.

OS WO200231134-A2.

PN 18-APR-2002.

PD 12-OCT-2001; 2001WO-US31874.

PF 12-OCT-2000; 2000US-240117P.

PR (FERR ) FERRING BV.

PA Qi S, Akinsanya KO, Riviere PJ, Junien J;

PI WPI: 2002-444178/47.

DR P-PSDB: ABG61596.

XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding the proteins, useful for treating e.g. fungal, bacterial, protozoan and viral infections, cancers, allergies, neurological disorders, or pain

PT -

XX Disclosure; Page 65-66; 113pp; English.

XX The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, CC psychotic and neurological disorders (e.g. anxiety, dementia, or CC schizophrenia), and dyskinesias. These may also be used in discovering CC therapeutic agents for the treatment of reproductive, inflammatory and CC metabolic disorders. ABK83322-ABK83343 encode human DPPP proteins.

XX SO Sequence 4829 BP; 1466 A; 886 C; 1017 G; 1460 T; 0 other.

Alignment Scores:

Pred. No.:	1.97e-245	Length:	4829
Score:	2402.00	Matches:	465
Percent Similarity:	82.01%	Conservative:	0
Best Local Similarity:	82.01%	Mismatches:	0
Query Match:	96.78%	Indels:	102
DB:	24	Gaps:	1

US-10-070-464-5 (1-465) x ABK83327 (1-4829)

QY 1 ThrGlyThraIaAsnProLyValIleThrPhelYseSerGluIleMetIleAspAlaGlu 20

DB 1165 ACAGGTRACGACAAATCTTAAGTCATCTTTTAAGATGTCAGAAATTAAGATTGATCTGAA 1224

QY 21 GlyArgIleIleAspValIleAspLyseGluLeuIleGlnProheGluIleAspGlu 40

DB 1225 GGAAGGATCATAGATGTCATAGATTAAGAACTAATTCACCTTTTGATTCATATTGAA 1284

QY 41 GlyValGluIuTyrtIleAlaAgaIaGlyTyrtThrProGluGlyLyseTyraIaTPSeIle 60

DB 1285 GGAGTTGAATATATGTCAGAGCTGAGATGACCTCGAGGAAATATATGCTTGCTCATC 1344

QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80

DB 1345 CTACTAGATGCTCCCAAGACTCCGCTACAGATAGTGTTCATCCACTGAATATTATTC 1404

QY 81 ProValGluAspAspValIleMetGluArgGluIleGluSerValProAspSerVal 100

DB 1405 CCACTAAGATAGATGTTATGAGAAAGCAGAGACTCTTAGTCAGTCCGATTCGTG 1464

QY 101 ThrProLeuIleIleTyrgIuGluThrThrAspIleTrpIleAsnIleHisAspIlePhe 120

DB 1466 ACCGCACTAATTCATGAGAAACACAGACATCGATTAATATCCATGACATCTTT 1524

QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCylys 140

DB 1525 CATGTTTTTCCCAAGATCAGAGAGAAATGAGTTATTTTGTCTCGAATGCAAA 1584

QY 141 ThrGlyPheArgHisLeuTyrtIleValIleThrSerIleLeuIleGluSerLyseTyrsArg 160

DB 1585 ACAGTTTCCGCTCATTTATACAAATTCATCTATTTAAGAGAAACAAATTAAGCA 1644

QY 161 SerSerGlyLyseLeuProAlaProSerAspPheLyseCyseProIleLyseGluIleAla 180

DB 1645 TCCAGTGTGGCTGCTGCTCCAGATGATTTCAAGTCTCTATCAAAAGAGAGATAGCA 1704

QY 181 IleThrSerGlyGluTyrtPgluValIleGluIleArgHisGlySerAsnIleGlnValAspGlu 200

DB 1705 ATTAACAGTGTGTAATGGAAGTTCTTGCCGCGATGATCTAATATCCAAAGTTGATGAA 1764

QY 201 ValArgArgLeuValTyrtPheGluGlyThrLyseAspSerProLeuGluHisIleLeuTyrt 220

DB 1765 GTCAGAAAGCTGTATTTTGAAGGACCAAGACCTCCCTTTAGCATCACTGTAC 1824

QY 221 ValIleSerTyrtValAsnProGlyGluValIleThrArgLeuThrAspArgIleTyrtSerHis 240

DB 1825 GTAGTCAGTACCTAATATCCGAGAGAGTACAAAGCTGACTACCTGCTACTACAT 1884

QY 241 SerCyseValIleSerGlnHisCyseAspPhePheIleSerTyrtSerAsnGlnLyseAsn 260

DB 1885 TCTTTCGTCATGCTGACGACGCTGACTCTTTTAATAGTAAGTATCAACGAAAGAT 1944

QY 261 ProHisCyseValSerLeuTyrtLyseLeuSerSerProGluAspAspProThrCyseTyrtThr 280

DB 1945 CCAACATGTGTGCTCTTTCAAGCATCAAGCTCGAAGATGACCCACTTGCAAAACA 2004

QY 281 LyseGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrtThrProPro 300

DB 2005 AAGGAATTTGGGCAACCATTTGGATTCAGCAGGTCCTCTCTGACTATACCTCCA 2064

QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrtGlyMetLeuTyrtLysePro 320

DB 2065 GAAATTTCTCTTTTGAAGATCACTGATTTTACATTTATGAGATGCTCTACAAAGCT 2124

QY 321 HisAspLeuGlnProGlyLyseTyrtProThrValLeuPheIleTyrgIyGly--ProG 340

DB 2125 CATGATCTACAGCTGGAAGAAATATCCATGCTGTGTCTCATATATGATGATGCTCTC 2184

QY 340 In----- 340

DB 2185 AGGTGAGTTGTGAATTAATCGGTTAAAGAGATCAAGTATTCGCTGAATACCTAG 2244

QY 340 ----- 340

DB 2245 CCTCTAGATTATGTGTTGTATGATAGACAAAGGGATCTGTCAAGGGCTTA 2304

QY 340 ----- 340

DB 2305 AATTGAAGCGCTTTAAATATAAATGAGGTCAAAATAGAAATTGACAGATCAGGTGAAG 2364

QY 340 ----- 340

DB 2365 GACTCAATATCTAGCTTCTCGATATGATTTGACTTAGATCGTGTGGGATCAGCAG 2424

QY 340 ----- 340

DB 2425 GCTGTCCTATGAGATACCTCTCCCTGATGAGCATTAATGACAGAGTCAATATCTTCA 2484

QY 341 --ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyArgThrGlyTrpThrg 360  
 Db 2485 GGGTGGTATTCGGGGGCCCACTCTCTGATCTTCTATGATACAGATACACGG 2544  
 QY 360 IuArgTyMetGlyHisProAspGlnAsnGlnGlyTyTrpLeuGlySerValAlaM 380  
 Db 2545 AACGTTATATGGGTCACCCGACAGATGAACAGGGCTATTACTTGAATCTGTGGCCA 2604  
 QY 380 etGlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuA 400  
 Db 2605 TGCAGAGGAAAGAGTCCCTGACCAATCGTTACTTACGTCTTACATGGTTCTCGG 2664  
 QY 400 spGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyL 420  
 Db 2665 ATGAGATATGTCATTTTGCACATACCAATATTTACTGATTTTATGAGGGCTGGAA 2724  
 QY 420 YspProTyArgPheLeuGlnIleTyProGlnGluArgHisSerIleArgValProGluSerG 440  
 Db 2725 AGCCATATGATTTTACATCTATCTCAGAGAGACACAGCATAGAGATTCTGAATCGG 2784  
 QY 440 IyGluHisTyArgLysLeuHisLeuLeuHisTyTrpLeuGlnGluAsnLeuGlySerArgIleA 460  
 Db 2785 GAGAAACATTAATGAACCTCATCTTTTGGACTTCAAGAAACCTTGATCAGATATTG 2844  
 QY 460 IAlaLeuLysValIle 465  
 Db 2845 CTCCTCTAAAAGTATA 2861  
 RESULT 15  
 AAD23843  
 ID AAD23843 standard; cDNA; 2510 BP.  
 XX  
 AC AAD23843;  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Human protease PRTS-2 cDNA.  
 XX  
 KM Human; protease; PRTS-2; tranquilliser; gene therapy; vaccine; allergy;  
 KM infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
 KM atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
 KM gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
 KM epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;  
 KM hypertension; neurological disorder; Parkinson's disease; drug screening;  
 KM cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;  
 KM diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;  
 KM autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
 KM developmental disorder; reproductive disorder; infertility; diarrhoea;  
 KM dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 616..2358  
 FT /\*tag= a  
 FT /product= "Human protease PRTS-2 protein"  
 XX  
 PN MO200183775-A2.  
 XX  
 PD 08-NOV-2001.  
 XX  
 PF 04-MAY-2001; 2001MO-US14651.  
 XX  
 PR 04-MAY-2000; 2000US-202082P.  
 PR 11-MAY-2000; 2000US-203566P.  
 PR 17-MAY-2000; 2000US-205803P.  
 PR 25-MAY-2000; 2000US-207477P.  
 PR 01-JUN-2000; 2000US-209402P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Deleageane AM, Lal P, Hafalia A, Patterson C, Walla NK, Kearney L,

PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;  
 PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM;  
 PI Reddy R, Yue H, Tang YT;  
 DR WPI; 2002-034518/04.  
 DR P-PSDB; AAE14337.  
 XX  
 PT Novel human proteases and polynucleotides encoding the proteases,  
 PT useful for treating, diagnosing or preventing cell proliferative,  
 PT cardiovascular, autoimmune/inflammatory, neurological and developmental  
 PT disorders -  
 PS Claim 5; Page 139-140; 151pp; English.  
 XX  
 CC The invention relates to human proteases (PRTS-14) and its corresponding  
 CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for  
 CC the diagnosis, treatment and prevention of disorders associated with  
 CC increased or decreased expression of PRTS. Examples of such disorders  
 CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
 CC hepatitis, psoriasis and cancer); autoimmune/inflammatory disorders  
 CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,  
 CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
 CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
 CC viral, bacterial, fungal, parasitic, protozoal and helminthic  
 CC infections); cardiovascular disorders (myocardial infarction, ischaemic  
 CC heart disease and hypertension); neurological disorders (epilepsy,  
 CC Alzheimer's disease, stroke, Huntington's disease, dementia,  
 CC Parkinson's disease, stroke, mental disorders including mood, anxiety  
 CC and seasonal affective disorder and prion diseases); gastrointestinal  
 CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);  
 CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,  
 CC alopecia, scabies, insect bites and urticaria); reproductive disorder  
 CC (infertility, disruption of estrous and menstrual cycle and  
 CC gynaecomastia); and developmental disorders (renal tubular acidosis,  
 CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
 CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic  
 CC fragments are useful for screening libraries of compounds in several drug  
 CC screening assays. The present sequence is human protease PRTS-2 cDNA.  
 XX  
 SO Sequence 2510 BP; 777 A; 494 C; 527 G; 712 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4,84e-229 Length: 2510  
 Score: 2244.50 Matches: 439  
 Percent Similarity: 84.03% Mismatches: 3  
 Best Local Similarity: 83.46% Indels: 5  
 Query Match: 90.43% Gaps: 6  
 DB: 24  
 US-10-070-464-5 (1-465) x AAD23843 (1-2510)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 822 ACAGGTACACCAATCTTAAGTCACCTTTAAGATGCAAGAAATTAAGATTGAGCTGAA 891  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 892 GGAAGATCATAGTGTCAATGATTAAGAACTAATTAACCTTTTGAGATTCTATTGAA 951  
 QY 41 GlyValGluTyTrpIleAlaArgAlaGlyTrpThrProGluGlyLysTyAlaTrpSerIle 60  
 Db 952 GAGTTGAATATATTCGACAGCTGAGTGACCTCTGAGGAAATATGCTTGTCATC 1011  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleLeuLeuSerProGluLeuPheIle 80  
 Db 1012 CTACTGATGCTCCACAGATCCCTCAAGATGATGTGTTGATCCACTGAAATATTATC 1071  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1072 CCAGTGAACATATATGTTATGAGAAAGCAAGACTCTTTAGTCAGGCTGATTCGTG 1131  
 QY 101 ThrProLeuIleIleTyArgGluGluThrThrAspIleTrpIleAsnIleHisAspIlePhe 120

```

Db      1132  ACGCCACTAATTATCTATGAAGAAACAAGACATCTGGATTAATATCCATGACATCTTT 1191
Qy      121   HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140
Db      1192  CATGTTTTTCCCAAGTCACGAAGAGAAATGAGTTATTTTGGCTTGAATGCAA 1251
Qy      141   ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160
Db      1252  ACAGGTTCCGTCATTTATACAAAATATCATCTATTTTAAAGAAACAAATATATAACA 1311
Qy      161   SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180
Db      1312  TCCAGTGGTGGGCTGCTGCTCCAACT----- 1338
Qy      181   IleThrSerGlyGluTyrGluVal-----LeuGly----- 190
Db      1339  GTCACT-----TGATGATGATCATTCATGAGATCTCTAGAGAACTCCATCTGTATG 1389
Qy      191   -----ArgHisGlySerAsnIleGluValAspGluValArgArgLeuValTyrPheGlu 208
Db      1390  TGTGTGACACATATATGTTAGATGCCAGTTCAGATGATGAGAGAGCTGTATATTTGAA 1449
Qy      209   GlyThrLysAspSerProLeuGluHisIleuTyrValIleSerTyrValAsnProGly 228
Db      1450  GGACACCAAGACTCCCTTTAGAGCATCCTGTAGTAGTACGTTACGTAATCTGGA 1509
Qy      229   GluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGlnHisCys 248
Db      1510  GAGGTGACAAAGCTGACGACCGTGTACTCATCTTGTGTCATCAGTCAGCATCTGT 1569
Qy      249   AspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyrLys 268
Db      1570  GACTTCTTTATAGAACTATAGTAACAGAAATCCACACTGTGTCTCCCTTACAAAG 1629
Qy      269   LeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIleLeu 288
Db      1630  CTATCAAGTCTTGAGATGACCCCACTGCANAAACAAAGAAATTTGGGCCACCAATTTG 1689
Qy      289   AspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSerThr 308
Db      1690  GATTGACGAGGCTCTTCTGACTATACTCTCCAGAAATTTCTCTTTGAAAGTACT 1749
Qy      309   ThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLys 328
Db      1750  ACTGATTTTACATGTATGAGATGCTCAACAGCCTCATGATCTACAGCTGGAAAGAAA 1809
Qy      329   TyrProThrValIleuPheIleTyrGlyGlyProGln----- 340
Db      1810  TATCTACTGTGCTGTTATATATGTGTGTCTCTCAGTGCACTGTGTAATATGCGTTT 1869
Qy      340   ----- 340
Db      1870  AAAGAGTCAAGTATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTGTGTAGTG 1929
Qy      340   ----- 340
Db      1930  ATAGACAAACAGGGGATCTGTACCGAGGGCTTAAATTGAAGCGCCTTAAATATAAA 1989
Qy      341   ---ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 359
Db      1990  ATGCTGTCTATGCTCGGGGCCCACTCACTGTGTGATCTTCTATGATACAGATACAG 2049
Qy      360   GluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 379
Db      2050  GAACGTTATATGGGTACCTGACCAAGATGAACAGGGCTATTACTTAGAGATCTGTGCC 2109
Qy      380   MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399
Db      2110  ATGCAGGCAAAAGTCCCTCTGACCAAAATCGTTTACATGAGTTCTCTG 2169
Qy      400   AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419
Db      2170  GATGAGAAATGTCATTGTCACATACAGATATATTACTGAGTTTATAGTGAAGGCTGGA 2229

```

```

Qy      420   LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
Db      2230  AAGCATATATGATTTA-----CAGGAGAGACACAGCATTAAGAGTTCTGGAATCG 2277
Qy      440   GlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIle 459
Db      2278  GGAGAACATTATATGAACCTGCATCTTTTGCACACTACCTCAAGAAAACCTTGATCACGTA 2337
Qy      460   AlaAlaLeuLysValIle 465
Db      2338  CCTGCTCTTAATAAGTGATA 2355

```

Search completed: October 15, 2003, 17:57:56  
Job time : 385.661 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 15, 2003, 17:29:02 Search time 72.8508 Seconds  
(without alignments)  
2817.309 Million cell updates/sec

Title: US-10-070-464-5  
Sequence: 1 TGTANPKVTFKMSIEMIDAE.....HLHYHQENIGSRIALKVI 465

## Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh  
-Q=/cg2\_1/USPTO.spool/US10070464/runat.15102003.113554.24845/app.query.fasta.1.2652  
-DB=Issued Patents NA -QFMT=fastag -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptio -NORM=none -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10070464 @CGN 1.1 146 @runat.15102003.113554.24845 -NCP1=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued Patents NA:\*  
1: /cg2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cg2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cg2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cg2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cg2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cg2\_6/prodata/1/ina/backfillseq.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	510.5	20.6	823	4	US-09-280-116-171
2	309.5	12.5	5496	4	US-09-462-284-1
3	293	11.8	612	4	US-09-392-184-31
4	238	9.6	3407	4	US-10-002-593-5
5	234	9.4	2924	5	PCT-US93-07923-1
6	214	8.6	2815	1	US-08-230-491A-1
7	214	8.6	2815	2	US-08-619-280A-1
8	214	8.6	2815	2	US-08-940-391-1
9	193	7.8	4982	3	US-08-699-103B-1
10	193	7.8	4982	4	US-09-229-059-1
11	193	7.8	4982	4	US-09-628-133-1
12	167	6.7	502	4	US-09-280-116-172

13	146.5	5.9	815	4	US-09-221-017B-962	Sequence 962, App
14	139	5.6	1896	3	US-09-016-080-2	Sequence 2, App1
15	132.5	5.3	543	4	US-09-221-017B-253	Sequence 253, App
16	124.5	5.0	657	4	US-09-221-017B-646	Sequence 646, App
17	110.5	4.5	3085	4	US-09-221-017B-1010	Sequence 1010, App
18	103.5	4.2	3942	3	US-09-162-484-19	Sequence 19, App1
19	103	4.1	904	4	US-09-221-017B-58	Sequence 58, App1
20	100.5	4.0	5798	2	US-08-658-665-178	Sequence 178, App
21	100.5	4.0	5798	3	US-08-796-101-39	Sequence 39, App1
22	100	4.0	1674	1	US-08-365-981-2	Sequence 2, App1
23	96.5	3.9	2424	1	US-08-821-119-16	Sequence 16, App1
24	95	3.8	679	4	US-09-221-017B-13	Sequence 13, App1
25	95	3.8	1664976	4	US-08-916-421B-1	Sequence 1, App1
26	94.5	3.8	7215	4	US-09-134-001C-627	Sequence 627, App
27	93	3.7	1230025	4	US-09-198-452A-1	Sequence 1, App1
28	92.5	3.7	535	4	US-09-389-681-428	Sequence 428, App
29	92.5	3.7	535	4	US-09-620-405B-428	Sequence 428, App
30	92.5	3.7	535	4	US-09-433-826B-428	Sequence 428, App
31	92.5	3.7	535	4	US-09-604-287A-428	Sequence 428, App
32	92.5	3.7	2406	1	US-08-396-479B-7	Sequence 7, App1
33	92.5	3.7	2406	1	US-08-818-823-7	Sequence 7, App1
34	92.5	3.7	2647	1	US-08-396-479B-9	Sequence 9, App1
35	92.5	3.7	2647	1	US-08-818-823-9	Sequence 9, App1
36	92.5	3.7	3969	1	US-08-396-479B-11	Sequence 11, App1
37	92.5	3.7	3969	1	US-08-818-823-11	Sequence 11, App1
38	92.5	3.7	4010	5	PCT-US94-07297-40	Sequence 40, App1
39	90.5	3.7	19446	4	US-08-261-527-51	Sequence 51, App1
40	92	3.7	2295	4	US-09-221-017B-889	Sequence 889, App
41	90.5	3.6	1404	4	US-09-134-001C-615	Sequence 615, App
42	90.5	3.6	5883	1	US-08-404-445-1	Sequence 1, App1
43	90	3.6	1416	1	US-08-236-311-3	Sequence 3, App1
44	90	3.6	1416	3	US-08-457-918-3	Sequence 3, App1
45	90	3.6	1508	1	US-08-236-311-6	Sequence 6, App1

## ALIGNMENTS

RESULT 1  
US-09-280-116-171  
Sequence 171, Application US/09280116A  
Patent No. 6331427  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs  
FILE REFERENCE: 5800-24, 035800/176965  
CURRENT APPLICATION NUMBER: US/09/280, 116A  
CURRENT FILING DATE: 1999-03-26  
NUMBER OF SEQ ID NOS: 268  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 171  
LENGTH: 823  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: prolyl oligopeptidases  
US-09-280-116-171

## Alignment Scores:

Pred. No.: 9.22e-51 Length: 823  
Score: 510.50 Matches: 115  
Percent Similarity: 85.71% Conservative: 5  
Best Local Similarity: 82.14% Mismatches: 5  
Query Match: 20.578 Indels: 15  
DB: 4 Gaps: 2  
US-10-070-464-5 (1-465) x US-09-280-116-171 (1-823)

QY 340 GINVAIAIAIAGLGYALProvalThrlauTrpIlePhetYrAspThrgIYrYr 359  
DB 151 AGGGTTCATTAGCTGGGCTCCAGACTCTGTGATCTTCATGATACAGATACAG 210  
QY 360 GIATGYrMeGlyHIsProAspGlnsngIugInGlyTYrTYrIeugIySerValaIA 379



```
Db 3815 GGTCCTTCACATTAGAGATCATACCGCCCTGTTTCTGACTGGCGTTTCTACGACTCA 3874
Qy 357 G|Y|T|Th|G|u|a|r|g|y|r|m|e|t|g|l|y|h|i|s|p|r|o|a|s|p|l|a|n|g|l|u|a|n|g|l|y|t|y|r|h|e|u|g|l|y 376
Db 3875 A|T|G|T|A|C|G|G|A|G|G|T|A|G|A|A|G|A|C|C|T|T|C|G|A|C|C|A|T|A|G|A|G|G|G|T|A|G|A|C|C|G|A|C|C 3934
Qy 377 S|e|r|v|a|l|a|w|e|t|g|l|a|a|g|l|u|y|s|h|e|p|r|o|s|e|r|g|l|u|p|r|o|a|n|a|g|l|e|u|e|u|h|i|s 396
Db 3935 G|C|G|T|C|---|C|G|A|A|G|A|C|G|G|G|T|T|C|A|G|A|C|G|T|G|A|G|G|G|G|A|T|T|C|T|T|G|A|T|C|G|A|C 3991
Qy 397 G|l|y|h|e|u|a|s|p|l|a|n|g|l|u|a|n|g|l|y|h|i|s|h|e|h|a|h|i|s|h|e|s|e|r|i|l|e|u|e|u|s|e|r|p|h|e|u|a|l 416
Db 3992 G|G|A|C|G|G|G|C|G|C|A|C|A|T|A|C|G|T|C|C|A|T|T|T|C|C|A|G|A|C|T|G|G|T|G|C|G|T|G|G|A|T|C|T|G|A|T 4051
Qy 417 A|r|g|a|l|a|g|l|y|s|h|e|p|r|o|t|y|r|a|s|p|l|e|u|a|n|g|l|y|t|y|r|p|r|o|g|l|u|a|n|g|l|y|h|i|s|h|e|s|e|r|i|l|e|u|a|l 436
Db 4052 G|G|G|G|A|T|G|C|---|G|T|C|T|C|T|G|A|G|A|G|T|C|C|A|T|T|G|C|A|T|G|G|A|T|G|G|T|T|C 4093
Qy 437 P|r|o|g|l|u|s|e|r|g|l|y|g|l|h|i|s|t|y|r|g|l|u|e|h|i|s 446
Db 4094 A|C|A|G|A|C|T|C|A|G|A|C|C|A|G|G|A|T|C|A|G|T|A|C|C|A|T 4123

RESULT 3
US-09-392-184-31/C
; Sequence 31, Application US/09392184
; Patent No. 6395869
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392,184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FaSTSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: prolyl oligopeptidase)
; NAME/KEY: misc_feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-392-184-31

Alignment Scores:
Pred. No.: 3,17e-25 Length: 612
Score: 293.00 Matches: 61
Percent Similarity: 66.35% Conservative: 8
Best Local Similarity: 58.65% Mismatches: 35
Query Match: 11.80% Indels: 0
Gaps: 0

US-10-070-464-5 (1-465) x US-09-392-184-31 (1-612)
Qy 340 G|l|v|a|l|a|l|e|a|a|g|l|a|p|r|o|v|a|l|t|h|r|e|u|t|r|i|l|e|h|e|t|y|r|a|s|p|h|r|g|l|y|t|y|r|h|e 359
Db 378 C|A|G|G|T|T|C|A|T|T|C|G|G|G|C|C|C|A|G|T|C|A|C|T|G|A|T|C|T|T|C|T|A|T|G|A|A|C|A|G|A|A|C|N|C|G 319
Qy 360 G|l|u|a|r|g|y|m|e|r|g|l|y|h|i|s|p|r|o|a|s|p|l|a|n|g|l|u|a|n|g|l|y|t|y|r|t|y|r|e|u|g|l|y|s|e|r|v|a|l|a 379
Db 318 G|G|A|C|G|T|T|A|T|G|G|T|C|A|C|C|T|N|N|C|A|G|N|A|T|G|A|C|A|G|G|G|C|T|A|T|T|A|C|T|T|G|A|T|G|G|C 259
Qy 380 M|e|t|g|l|a|l|g|l|u|a|s|h|e|p|r|o|s|e|r|g|l|u|p|r|o|a|n|a|g|l|e|u|e|u|h|i|s|g|l|y|h|e|u 399
Db 258 A|T|T|C|C|A|N|C|A|G|A|A|G|T|T|C|C|C|T|C|T|G|A|N|C|C|A|T|N|N|G|T|N|C|T|T|C|T|T|A|N|A|T|G|G|T|T|C|G 199
Qy 400 A|s|p|g|l|u|a|n|v|a|h|i|s|h|e|h|a|h|i|s|h|e|s|e|r|i|l|e|u|e|u|s|e|r|p|h|e|u|a|l|a|g|l|y 419
Db 198 G|A|T|T|A|G|A|T|T|C|C|A|T|T|T|C|A|N|A|C|C|A|G|A|N|A|N|A|T|A|C|T|T|A|G|G|T|T|T|A|G|G|G|C|T|G|G 139
```

```
Qy 420 l|y|e|p|r|o|t|y|r|a|s|p|l|e|u|a|n|g|l|y|t|y|r|p|r|o|g|l|u|a|n|g|l|y|h|i|s|h|e|s|e|r|i|l|e|u|a|l|p|r|o|g|l|u|s|e|r 439
Db 138 A|A|C|C|A|A|T|G|A|T|T|N|C|A|A|C|C|A|C|T|C|N|G|A|A|G|A|N|C|C|A|C|A|N|A|G|G|T|N|C|T|G|A|T|C|G 79
Qy 440 G|l|y|l|u|h|i|s|t|y|r 443
Db 78 G|G|A|G|A|C|A|A|T|A|T 67

RESULT 4
US-10-002-593-5
; Sequence 5, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTI
; FILE REFERENCE: Acty Docket No. 6586198 1242/46/2
; CURRENT APPLICATION NUMBER: US/10/002,593
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/244,524
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3407
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-593-5

Alignment Scores:
Pred. No.: 1.75e-17 Length: 3407
Score: 238.00 Matches: 120
Percent Similarity: 33.51% Conservative: 71
Best Local Similarity: 21.05% Mismatches: 193
Query Match: 9.59% Indels: 186
Gaps: 24

US-10-070-464-5 (1-465) x US-10-002-593-5 (1-3407)
Qy 2 G|l|y|h|r|a|l|a|s|p|r|o|l|y|s|v|a|l|t|h|r|p|h|e|y|m|e|r|s|e|r|g|l|u|e|t|i|l|e|a|p|a|l|a|g|l|y 21
Db 853 G|G|A|G|C|T|G|T|G|A|T|C|A|C|T|G|T|A|A|G|T|C|T|T|G|T|---|G|T|A|A|T|A|C|A|G|C|T|C|T|C|A|G 906
Qy 22 A|r|g|i|l|e|l|e|a|p|v|a|l|l|e|a|p|l|y|s|g|l|u|e|h|i|e|g|i|n|p|r|h|e|g|i|u|l|e|u|h|e|u|p|h|e|g|l|y 41
Db 907 T|C|A|G|T|C|A|C|C|A|T|G|C|A|C|T|T|C|A|T|C|A|A|T|C|A|T|G|C|T|G|C|T|T|C|T|A|T|G|T|G|A|T|A|---|G|G 963
Qy 42 V|a|l|g|l|u|t|r|i|l|e|a|l|a|r|g|a|l|a|g|l|y|t|p|h|r|p|r|o|g|l|u|g|l|y|l|y|s|t|y|r|a|l|a|t|r|p|s|e|r|i|l|e|u 61
Db 964 G|A|T|C|A|C|T|A|C|T|T|G|T|G|A|T|G|T|G|A|C|A|T|G|G|C|A|---| 993
Qy 62 L|e|u|a|s|p|a|r|s|e|r|g|i|h|r|a|r|g|l|e|u|g|i|h|i|l|e|a|l|l|e|u|h|i|s|e|r|p|r|o|g|l|u|e|h|e|u|h|i|p|r|o 81
Db 994 ---|A|C|A|C|A|A|A|A|G|A|T|T|C|T|T|G|C|A|G|T|G|C|---| 1023
Qy 82 V|a|l|g|l|u|a|s|p|a|r|v|a|l|e|r|g|l|u|a|r|g|l|a|r|g|l|e|h|i|e|g|l|u|s|e|r|v|a|l|p|r|o|a|s|p|s|e|r|v|a|l|t|h|r 101
Db 1024 ---|A|G|A|G|A|T|T|C|A|G|A|C|A|T|A|T|---|T|C|G|T|C|A|G 1050
Qy 102 P|r|o|l|e|h|i|l|e|l|y|r|g|l|u|n|t|h|r|h|a|r|p|i|l|e|r|i|l|e|a|s|n|i|l|h|i|s|a|p|i|l|e|h|i|s 121
Db 1051 G|A|T|A|T|T|G|A|C|A|T|A|G|A|T|A|T|C|G|T|G|A|G|A|T|G|---|A|A|C|G|C|---| 1092
Qy 122 V|a|l|p|h|e|r|p|r|o|g|l|u|s|e|r|i|s|g|l|u|g|i|h|i|l|e|g|i|h|i|l|e|h|a|h|i|s|e|r|g|l|u|c|y|s|t|y|r|h 141
Db 1093 T|T|A|G|T|G|C|A|C|G|G|C|A|C|A|C|A|T|T|G|A|A|T|G|A|T|A|C|T|A|C|G|G|C|T|G|G|T|T|G|G|A|A|G|A|T|T|A|G|C|T 1152
Qy 142 G|l|y|h|e|a|r|h|i|s|---|L|e|u|T|y|l|e|h|i|s|h|e|s|e|r|i|l|e|u|h|i|s 154
Db 1153 T|C|A|G|A|C|T|C|A|T|T|A|C|C|T|T|G|A|T|G|T|A|T|A|C|T|T|C|T|A|C|A|G|A|T|A|T|C|A|G|C|---|A|A|T 1206
```

QY 155 GluSerIysTyrLysArgSerSerGlyLeuProAlaProSerAspPheLysCysPro 174  
 DB 1207 GAAGAGGTTACGACACAT-----TCGATTCCAAATAGAT 1245  
 QY 175 IleuSgluIleAlaIleThrSerGlyuTyrPgluValLeuGlyArgHisGlySer 194  
 DB 1246 AAAAAAGCTCAGATTTATTACAAAGGACCTGGGAAGTCATCGG----- 1293  
 QY 195 AsnIleGlnValAspGluValArgArgLeuValTyrPheGluGlyThr---LysAspSer 213  
 DB 1294 ---ATGAAGCTTACACAGATGATTATCTATACCTAGTAAATGATATTAAGATG 1350  
 QY 214 ProLeuGlnHisIleuTyrValValSerTyrValAsnProGlyGluValThrArgLeu 233  
 DB 1351 CCAGAGAGAGAAATCTTATTAATAACCACTTATGACTTACAAAAGTACATGCTTC 1410  
 QY 234 ThrAspArgGlyTyrSerHisSerCysGlySer---GlnHisCysAspPhePheIle 252  
 DB 1411 -----AGTGTAGCTGAATCCGAAAGGTGCTAGTACTATCT 1449  
 QY 253 SerIysTyrSerAsnGlnLysAsnProHis-----Cys-----ValSer 265  
 DB 1450 GTGTGATTCAGTAAAGAGCGCAAGATATATACGTGAGATGTCGGTCTGCTGCC 1509  
 QY 266 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrrPala 285  
 DB 1510 CTCTATACTCTACACAGCGCTGATATGAAGGCTGAGAGCTCTGGAAC---AAT 1566  
 QY 286 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGlu-----Ile 302  
 DB 1567 TCAGCTTGGAAATAATG-----CTGCAGATGTCCAGATGCCCTCCAAAAAAGTGCAC 1620  
 QY 303 PheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAsp 322  
 DB 1621 TTCATTATTGGAATGAAGAAATAATTT---TCGTATCAGATGATCTCTCCCTCAT--- 1674  
 QY 323 LeuGlnProGlyLysLeuTyrProThrValLeuPheIleTyrGlyGlyPro----- 339  
 DB 1675 TTTGATTAATCCAGAAATATCTCTACTATTAGATGTGTAGCAGGCCCATGTGCA 1734  
 QY 339 ----- 339  
 DB 1735 AAAGCAGACACTGCTTTCAGACTGAAGTGGGCACTTACCTTGCAAGCAGAAAAAATT 1794  
 QY 339 ----- 339  
 DB 1795 ATAGTAGCTAGCTTGTATGCGAGAGAAAGTGGTTACCAAGAGATAGATCATGCATGCA 1854  
 QY 339 ----- 339  
 DB 1855 ATCAACAGAGACTGGGAACATTTGAAGTGAAGATCAAAATTGAAGCCAGACAAATT 1914  
 QY 339 ----- 339  
 DB 1915 TCMAAAATGGGATTTGTGGAACAACAAGAAATTGCAATTTGGGGCTGGTCATATGAGGG 1974  
 QY 340 -----GlnValAlaIleAlaGly 345  
 DB 1975 TACGTAACTCAATGCTCTGGGATCGGGAAGTGGCGTGTTCAGATGTGGAATAGCCGTG 2034  
 QY 346 AlaProValThrLeuTrrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGly--- 364  
 DB 2035 GCGCTGTATCCGGGTGGAGTACTATGACTAGTGTACAGAACGTTACTATGAGGTCTC 2094  
 QY 365 ---HisProAspGlnAsnGlnGlnIleGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 383  
 DB 2095 CCAACTCCAGAAAGACCTTGACCATTAACAAGAAATTCAAGCATGATGAGCACTGAA 2154  
 QY 384 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 403  
 DB 2155 AATTT-----AACCAAGTTGAGTACCTCTTATTCATGAGAACAGCATGATACGTT 2208  
 QY 404 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 423

DB 2209 CACTTACAGCAGTAGCTCAGATCTCCAAAGCCCTGGTTCAGTGTGGATGATTTCCAG 2268  
 QY 424 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyLysIleTyr 443  
 DB 2269 GCATGTGCTTACTGATGAAAGACCATGGAATACTGACGACAGACACCAACATATTA 2328  
 QY 444 GluLeuHisLeuLeuHisTyrLeuGlnGlu 453  
 DB 2329 TATACCAATGAGCCCACTTCATTAACAA 2358  
 RESULT 5 0 4  
 PCT-US93-07923-1  
 ; Sequence 1, Application PC/TUS9307923  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Morimoto, Chikao  
 ; APPLICANT: Schlossman, Stuart F.  
 ; APPLICANT: Tanaka, Tohichiaki  
 ; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; COMPUTER: IBM PS/2 Model 502 or 55SX  
 ; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
 ; SOFTWARE: WordPerfect (Version 5.0)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/07923  
 ; FILING DATE: 19930819  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/934,162  
 ; FILING DATE: 21-AUG-1992  
 ; APPLICATION NUMBER: 07/832,211  
 ; FILING DATE: 06-FEB-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Frazer, Janis K.  
 ; REGISTRATION NUMBER: 34,819  
 ; REFERENCE/DOCKET NUMBER: 00530/055002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 542-5070  
 ; TELEFAX: (617) 542-8906  
 ; TELEX: 200154  
 ; INFORMATION FOR GEO ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2924  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; PCT-US93-07923-1  
 Alignment Scores:  
 Pred. No.: 4,05e-17 Length: 2924  
 Score: 234.00 Matches: 120  
 Percent Similarity: 33.10% Conservative: 70  
 Best Local Similarity: 20.91% Mismatches: 190  
 Query Match: 9.43% Indels: 194  
 DB: 5 Gaps: 25  
 US-10-070-464-5 (1-465) x PCT-US93-07923-1 (1-2924)  
 QY 2 GlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGly 21  
 DB 788 GGAAGCTGGAATCCAACTGTAAGTCTTTGTT-----GTAATATACGACTCTCTCAGC 841  
 QY 22 ArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGly 41



```

Db      842 TCAGTCACCAATGCAATTCATCAATCACTGCTCTGCTTATGTTGATA---GGG 898
Qy      42 ValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTrpSerIleLeu 61
Db      899 GATCACTACTGTGTGATGTGATGCATGGCA----- 928
Qy      62 LeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIlePro 81
Db      929 -----ACACAAGAAAGAAATTTCTTGGAGTGGCTC----- 958
Qy      82 ValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThr 101
Db      959 -----AGGAGGATTCAGAACTAT-----TCGGTCAATG 985
Qy      102 ProeuIleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHis 121
Db      986 GATATTGTGACTGATGATGAATCCAGTGAAGATGG---AACTGC----- 1027
Qy      122 ValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThr 141
Db      1028 TTAGTGGCAGCGCAACATGTAATGATGACTAGCTGGCTGGTTGGAAGATTTAGGCTT 1087
Qy      142 GlyPheArgHis-----LeuTyrLysIleThrSerIleLeuLys 154
Db      1088 TCAGAACCTCATTTTACCCCTTGATGTAAATAGCTTTCACAGATCATCAGC-----AAT 1141
Qy      155 GluSerLysTyrLysArgSerSerGlyLysLeuProAlaProSerAspPheLysCysPro 174
Db      1142 GAAGAACGTTACACACACATT-----TCCTATTTCGAATATGAT 1180
Qy      175 IleGluGluGluIleAlaIleThrSerGlyGluTrpGluValLeuGlyArgHisGlySer 194
Db      1181 AAAAAAGACTGCACATTATTATTAACAAGACCTGGGAAGTCATCGGG----- 1228
Qy      195 AsnIleGlnValAspGluValArgArgLeuValTyrPheGluGlyThr---LysAspSer 213
Db      1229 ---ATAAGAGCTTAACCAAGTATATCTATACATTAAGTAAATGAAATAAAGAAATG 1285
Qy      214 ProeuGluHisIleLeuTyrValIleValSerTyrValAsnProGlyGluValThrArgLeu 233
Db      1286 CCAGAGAGAAAGAGATCTTTATAAATC-----CAACTT 1318
Qy      234 ThrAspArgGlyTyrSerHisSerCysCysIleSer-----GlnHisCys 248
Db      1319 AGTGAC-----TATACAAAAGTGAACATGCCTCAGTTGTGAGTGAATCCGGAAGAGTGT 1372
Qy      249 AspPheHeIleSerLysTyrSerAsnGlnLysAsnProHis-----Cys----- 263
Db      1373 CAGTACTATTCTGTGTCATTCAAGTAAAGCGCAAGTATATCAGCTGAGATGTTCCGCT 1432
Qy      264 -----ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLys 281
Db      1433 CTGGTGTGCCCCCTTACTTACTTACACAGCAGCTGATGATTAAGGCGTGAAGCTCTG 1492
Qy      282 GluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGlu 301
Db      1493 GAAGAC---AATTCAGCTTGGATTAATG---CTGCAGAAATGCGAATGCCCTCC 1543
Qy      302 -----IlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 318
Db      1544 AAAAACTGAGCTCATTAATTTGAAAGAAACAAATTT---TGTATCAGATGATCTTG 1600
Qy      319 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly 338
Db      1601 CCTCCTCAT---TTGATTAATCCAGAAATATCCTTACTATTAAGTGTGATGAGCG 1657
Qy      339 Pro----- 339
Db      1658 CCATGTAGTCAAAAAGACAGACTGTCTCAGACTGAATGGGCACTTACCTTGCAAGC 1717
Qy      339 ----- 339
Db      1718 ACAGAAAAACATTATAGTACTGCTTGTGATGCAAGAAAGTGTTACCAAGAGATTAAG 1777

```

```

Qy      339 ----- 339
Db      1778 ATCATGATGCAATCAACAGAAAGCTGGAACTTTGAAGTTGAAGATCAAAATGAAGCA 1837
Qy      339 ----- 339
Db      1838 GCCAGACAATTTTCAAAAATGGGATTTGTGACAAACAAACGAATTGCAATTTGGGGCTGG 1897
Qy      340 -----GlnVal 341
Db      1898 TCATATGAGGGGTACGTAACTCAATGGTCTGGGATGACGAAGTGGCTGTTCAAGTGT 1957
Qy      342 AlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspTrpGlyTyrThrGluArg 361
Db      1958 GGAATAGCCGTGGCGCGCTGTAATCCCGGGGAGTACTATGACTCAGTGTACACAAAGCT 2017
Qy      362 TyrMetGly-----HisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 379
Db      2018 TACATGGGTCTCCCACTCCGAAAGACAACTTGACCTTACAGAAATTCACACAGTCAATG 2077
Qy      380 MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399
Db      2078 AGCAGACTGAAAATTTT-----AAACAAGTGAAGTACCTCTTATTCATGGAACAGCA 2131
Qy      400 AspGluMetValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419
Db      2132 GATATTAACGTTCACTTACAGCACTGACATGATCTCCAAAGCCCTGCTGATGTTTGA 2191
Qy      420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
Db      2192 GTGATTTCCAGGCAATGTGTATCTGATGAAACCAATGGAATTAAGCTAGACAGCACAGCA 2251
Qy      440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlu 453
Db      2252 CACCAACATATATATACCCACATGAGCAGCACTTATAAACA 2293

RESULT 6
US-08-230-491A-1
; Sequence 1, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FELFE & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 2815 Base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
-08-230-491A-1

Alignment Scores:  
Seq. No.: 8,88e-15 Length: 2815  
Score: 214.00 Matches: 87  
Percent Similarity: 32.48% Conservative: 53  
Percent Local Similarity: 20.19% Mismatches: 127  
Percent Local Similarity: 8.62% Indels: 164  
Query Match: 1 Gaps: 13

S-10-070-464-5 (1-465) x US-08-230-491A-1 (1-2815)  
Y 140 LysThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159  
b 1319 AAGGATGGCTACAAACATATCTACTATATC----- 1348  
y 160 ArgSerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluLe 179  
b 1349 -----AAAGCACTGTGAAATATCTATT 1372  
y 180 AlaIleThrSerGlyGluTyrPheGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199  
b 1373 CAATTTACAGTGGCAAGTGGAGGCCATA-----AATATA----- 1408  
y 200 GluValArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeu 219  
b 1409 -----TTCAAGATACACAGGATTCCTGTTTATTCTAGCAAT 1447  
y 220 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgLysTyrSer 239  
b 1448 GAATTTGAAGATACCTCGAAGAGAAACATCTACAGATTAGCACTTGGAACTTCTCT 1507  
y 240 HisSer---CysCysIleSerGlnHis-----CysAspPhePheIleSer 253  
b 1508 CCAGGCAAGAGAGTGTACTTCTGCACTTACAGAAAGAAAGGTCGCAATTTACACAGCA 1567  
y 254 LysTyrSerArgGlnLysAsnProHisCysValSerLeuTyr-----LysLeu 269  
b 1568 AGTTTCAGCGCTACGCCCACTATGCACTTGTCTGCTACGCCCAAGCATCCCACTT 1627  
y 270 SerSerProGluAspAspProThrCysLysThrLysGluPheIlePalaThrIleLeuAsp 289  
b 1628 TCACCCCTTCATGATGACGACATGATCAAGAA-----ATTAAATCTCTGAA 1675  
y 290 SerAlaGlyProLeuProAsp-----TyrThrProProGluIlePheSer 304  
b 1676 GAAACCAAGGATTTGAAATATCTTTGAAATAATTCACAGCTGCTAAAGAGAAATTAG 1735  
y 305 PheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGln 324  
b 1736 AAACCTGAAGTAGATGAATTAATCTTTATGTAACAAGATGATCTTCTCTCTCAATTGAC 1795  
y 325 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro----- 339  
b 1796 AGATCAAGAGATATCCCTTCTTAATTCATGATGATGATGATGATGATGATGATGAT 1855  
y 339 ----- 339  
b 1856 AGCTGTGATTTCTGTTAATGATATCTTCTGCAAGTAAGAAAGGATGAT 1915  
y 339 ----- 339  
b 1916 GCCTGTGATGATGTCGAGAAACAGCTTCCAAAGTGACAAACCTCTATGACGTAT 1975  
y 339 ----- 339  
b 1976 CGAAAGCTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2035  
y 339 ----- 339

Db 2036 ATGGGTTTCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2095  
y 340 -----GlnValAlaIleAlaGlyAlaPro 347  
Db 2096 TCATCACTGGCCCTTGATCTGGAAGTGTCTTTTCAAATGTGATATGACAGTGGCTTCA 2155  
y 348 ValThrLeuTyrPheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro--- 366  
Db 2156 GTCTCCAGCTGCGAATATATACCGCTCTCTCTACACAGAGATATCATGTCCTCCCAACA 2215  
y 367 ---AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhe 385  
Db 2216 AAGATGATATATCTTGACACACTAAGATTAATCACTGTGATGCGCAAGACCAATATTC 2275  
y 386 ProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 405  
Db 2276 AGAAATGTAGAC-----TATCTTTCATCCACGGAACAGCATATGATATGACACTT 2329  
y 406 AlaIleThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 425  
Db 2330 CAAACTCAGCACAGATGTGCTAAAGCTCTGTATATGACCAAGTGAATTTCCAGGCAATG 2389  
y 426 IleTyrProGlnGluArgHisSerIleArgValProGluSerGly-----GluHis 442  
Db 2390 TGGTACTCTGACCAAGAACACACGCGCTTA-----TCGGCTCTGTCACAGAACCCAC 2437  
y 443 TyrGluLeuHisLeuLeuHisTyrLeuGlnGlu 453  
Db 2438 TTATACACCCACATGACCACTCTCTAAAGCAG 2470

RESULT 7  
US-08-619-280A-1  
Sequence 1, Application US/08619280A  
Patent No. 5767242  
GENERAL INFORMATION:  
APPLICANT: Zimmermann, Rainier, Park, John E.;  
TITLE OF INVENTION: RETIG, Wolfgang; Old, Lloyd J.  
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pelfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
COMPUTER: IBM PS/2 PC-DOS  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/619,280A  
FILING DATE: 18-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/230,491  
FILING DATE: 20-APRIL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5767242man D.  
REGISTRATION NUMBER: LUD 5330.1  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 638-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2815 Base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-619-280A-1

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
8	886.15	214.00	32.48%	20.19%	2815	87	53	127	164	13

US-10-070-464-5 (1-465) x US-08-619-280A-1 (1-2815)

```

QY 140 LysThrGlyPheArgHisLeuTyrLysLeuHisSerIleLeuYsgIuSerLysTyrLys 159
DB 1319 AAGCATGGCTACAAACATATTCATCATATC----- 1348
QY 160 ArgSerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGluIle 179
DB 1349 -----AAAGACACTGTGAAATGCTATT 1372
QY 180 AlaIleThrSerGlyIuTProGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199
DB 1373 CAATTTCAGTGGCAAGTGGAGGCCATA-----AAATATA----- 1408
QY 200 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeu 219
DB 1409 -----TTCAGAGTAAACACAGATTCACTGTTTATTCTAGCAAT 1447
QY 220 TyrValValSerTyrValAlaenProGlyGluValIleThrArgLeuThrAspArgLysTyrSer 239
DB 1448 GAATTTGAAAGAAATCCCTGGAGAGAAACATCTACAGATTGACATTGGAGCTATCCCT 1507
QY 240 HisSer---CysCysIleSerGlnHis-----CysAspPhePheIleSer 253
DB 1508 CCAAGCAAGAGAGTGTGTACTTCCCATCTTAGAAGAAAGTGGCAATATTTACACAGCA 1567
QY 254 LysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyr-----LysLeu 269
DB 1568 AGTTTCAGCGACTACCGCAAGTACTATGCACTGTCTGCTACGCGCCACAGCATCCCAT 1627
QY 270 SerSerProGluAspAspProThrCysLysThrLysGluPheIleAlaThrIleLeuAsp 289
DB 1628 TCCACCTTCATGATGAGCGACTGATCAGAA-----ATTAATATCCCTGGAA 1675
QY 290 SerAlaGlyProLeuProAsp-----TyrThrProProGluIlePheSer 304
DB 1676 GAAACCAAGCAATTTGAAATATCTTTGAAATATATCAGCTGCTTAAAGGCAATTAAG 1735
QY 305 PheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGln 324
DB 1736 AAAGTGAAGTAGATGAATTAATCTTATGATGTAACAAGATGATCTTCTCTCAATTGAC 1795
QY 325 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro----- 339
DB 1796 AGATCAAAAGATATCCCTGCTTAATTAAGTGTATGATGATGATGATGATGATGATGAT 1855
QY 339 ----- 339
DB 1856 AGCTCTGATTTGCTCTTAATTGATATCTTATCTTCAAGTAGAAGAGAGATGTCATT 1915
QY 339 ----- 339
DB 1916 GCCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1975
QY 339 ----- 339
DB 1976 CGAAAGCTGGTGTATTGAAAGTGAAGACAGATTACAGCTGTGAGAAATTCATAGAA 2035
QY 339 ----- 339
DB 2036 ATGGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2095
QY 340 -----GlnValAlaIleAlaGlyAlaPro 347
  
```

```

DB 2096 TCATCACTGGCCCTTCGATCTGGAACGTGCTTTTCAAAATGTGATATAGCATGCTTCA 2155
QY 348 ValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro--- 366
DB 2156 GTCTCCAGCTGGGAATATTACGGCTGTCTGTACACAGAGATTCATGGCTGCCAACAA 2215
QY 367 ---AspGlnMetGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhe 385
DB 2216 AAGGATGATATATCTTGTAGCACTATAGAAATTCAACTGTGTGGCAAGAGCAATATTTTC 2275
QY 386 ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 405
DB 2276 AGAAATTAAC-----TATCTTCTCATCTACAGAACAGACAGATGATTAATGTCACCTTT 2329
QY 406 AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 425
DB 2330 CAAAACTCACACAGATGTCTAAAGCTGTGTTAATCACAAGATTCCTCCAGGCAATG 2389
QY 426 IleTyrProGlnIleArgHisSerIleArgValProGluSerGly-----GluHis 442
DB 2390 TGGTACTCTGACCAAGAACCAAGGCTTA-----TCCGCGCTGTCCAGAACCAAC 2437
QY 443 TyrGluLeuHisLeuLeuHisTyrLeuGlnGlu 453
DB 2438 TTATACACCAATGACCACTTCTTAAGCAG 2470

RESULT 8
US-08-940-391-1
; Sequence 1, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer, Park, John E.;
; APPLICANT: Reltig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940.391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 596537man D.
; REGISTRATION NUMBER: 30,946
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-940-391-1
  
```

Alignment Scores:  
 Pred. No.: 8.88e-15 Length: 2815  
 Score: 214.00 Matches: 87  
 Percent Similarity: 32.48% Conservative: 53  
 Best Local Similarity: 20.19% Mismatches: 127  
 Query Match: 8.62% Indels: 164  
 DB: Gaps: 13

US-10-070-464-5 (1-465) x US-08-940-391-1 (1-2815)

140 LysThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159  
 1319 AAGGATGGCTACCAAGATATTCCTACTATATC----- 1348  
 160 ArgSerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluIle 179  
 1349 -----AAAGACACTGTGCAAAATGCTATT 1372  
 180 AlaIleThrSerGlyLysTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199  
 1373 CAATATTACAAAGTGGCAAGTGGAGGCCATA-----AATATA----- 1408  
 200 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeu 219  
 1409 -----TTCAGAGTAAACAGAGATTCACTGTTTATTCTTACCAAT 1447  
 220 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgLysTyrSer 239  
 1448 GAATTGGAAGATATACCTCGAAGAAAGAAACATCTACAGAAATTGCAATGGAGCTATCC 1507  
 240 HisSer---CysCysIleSerGlnHis-----CysAspPhePheIleSer 253  
 1508 CCAAGGAAGAAAGTGTACTTCTGCTTAAGAAAGAAAGGCGCAATATTACACAGCA 1567  
 254 LysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyr-----LysLeu 269  
 1568 AGTTTACAGCAGCAACCCCAAGTACTATGCCTGTCTGACAGCCCAAGCACCCTATT 1627  
 270 SerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThrIleLeuAsp 289  
 1628 TCCACCTTCATGATGAGCAGCAGCTCAAGAA-----ATTAAATCCCTGGA 1675  
 290 SerAlaGlyProLeuProAsp-----TyrThrProGluIlePheSer 304  
 1676 GAAACAAAGAAATTGAAATGCTTTGAAAAATATCCAGCTGCTAAAGGGAATTAAG 1735  
 305 PheGluSerThrThrGlyPheThrLeuTyrIleMetLeuTyrLysProHisAspLeuGln 324  
 1736 AAACCTGAAGTATGATGAATTAATTAATGTTATGTTACAAAGATGATCTCTCTCAATTGAC 1795  
 325 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro----- 339  
 1796 AGATCAAGAAAGTATCCCTGCTTAATTAATCAAGTATGTTGTCCTGACAGTCAAGTGA 1855  
 339 ----- 339  
 1856 AGGCTGTATTGCTGTTAATTGATATCTTATTCTTGCAAGTAAGAAAGGATGTCATT 1915  
 339 ----- 339  
 1916 GCCTTGTGATGTGTGAGGAAACAGCTTCCAAAGTGACAAACTCCTTATGCAAGTAT 1975  
 339 ----- 339  
 1976 CGAAGCTGGTGTATTATGAAGTTGAAGACAGATTACAGCTGCAAAAATTCAATGA 2035  
 339 ----- 339  
 2036 ATGGGTTTCAATTGATGAAAAAAGATAGCATATGGGGCTGCTCTATGAGATATCGTT 2095  
 340 -----GlnValAlaIleAlaGlyAlaPro 347  
 2096 TCATCACTGACCCTTGATCTGAACTGCTTTTCAAAATGTGTATAGCAGTGGCTCCA 2155

348 ValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro--- 366  
 2156 GTCCTCAGCTGGGAATATTACCGCTGTCTTACACAGAGATTCATGGGCTCCCAACA 2215  
 367 ---AspGlnAsnGluGlnIleTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhe 385  
 2216 AAGGATGATATCTTGAGCATATTAAGAAATTCATGTAATGGCAAGACAGATATTTC 2275  
 386 ProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 405  
 2276 AGAAATGTAGAC-----TATCTTCCATCCACGAAACAGCATATATATGCACTTT 2329  
 406 AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 425  
 2330 CAAACTCAGACAGATGCTCTAAGCTGTGTTAATGCAAGTGAATTCAGGCAATG 2389  
 426 IleTyrProGlnGluArgHisSerIleArgValProGluSerGly-----GluHis 442  
 2390 TGGTACTTGACCAAGAACAGGCTTA-----TCCGGCTGTCCACGAACAC 2437  
 443 TyrGluLeuHisLeuLeuHisTyrLysGluGlu 453  
 2438 TTATACACCCACATGACCATTCTTAAGCAG 2470

RESULT 9  
 US-08-699-103B-1/c  
 ; Sequence 1, Application US/08699103B  
 ; Patent No. 6107462  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rine, Jaaper D.  
 ; APPLICANT: Hampton, Randolph  
 ; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING  
 ; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Fish & Richardson P.C.  
 ; STREET: 2200 Sand Hill Road, Suite 100  
 ; CITY: Menlo Park  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94025  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/699,103B  
 ; FILING DATE: 16-AUG-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/002,581  
 ; FILING DATE: 17-AUG-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Green, Grant D.  
 ; REGISTRATION NUMBER: 31,259  
 ; REFERENCE/DOCKET NUMBER: 09272/005001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650/322-5070  
 ; TELEFAX: 650/854-0875  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4982 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; US-08-699-103B-1  
 Alignment Scores:  
 Pred. No.: 6.98e-12 Length: 4982  
 Score: 193.00 Matches: 65  
 Percent Similarity: 40.97% Conservative: 28



Qy 390 AsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAenValHisPheAlaHisThrSer 409  
Db 4470 AATGATTTTGTGATCGACGAAACAGAGATGATTAACGTTCAATTCACAAATTCCTTA 4411  
Qy 410 IleLeuLeuSerPheLeuValArgAlaGly--LysProTyrAspLeuGlnIleTyrPro 428  
Db 4410 AAGTTTCTGGACCTTTTGATCTTAATGTTGTGAAATTTATGACGTCACGCTCTTCTCT 4351  
Qy 429 GlnGluArgHisSerIleArg 435  
Db 4350 GACTCAGATCATAGTATAGA 4330

RESULT 11  
US-09-628-133-1/c  
; Sequence 1, Application US/09628133  
; Patent No. 6531292  
; GENERAL INFORMATION:  
; APPLICANT: Rine, Jasper D.  
; APPLICANT: Hampton, Randolph  
; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING  
; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 2200 Sand Hill Road, Suite 100  
; CITY: Menlo Park  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/628.133  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/699,103  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Grant D.  
; REGISTRATION NUMBER: 31,259  
; REFERENCE/DOCKET NUMBER: 09272/005001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/322-5070  
; TELEFAX: 650/854-0875  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4982 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; US-09-628-133-1

Alignment Scores:  
Pred. No.: 6.98e-12 Length: 4982  
Score: 193.00 Matches: 65  
Percent Similarity: 40.97% Conservative: 28  
Best Local Similarity: 28.63% Mismatches: 80  
Query Match: 7.78% Indels: 54  
Gaps: 9

US-10-070-464-5 (1-465) x US-09-628-133-1 (1-4982)

Qy 244 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAenGlnLysAenProHisCys 263  
Db 4953 TTAAGTGCACCTATCTCTGTTTCTTTGCAATGGGGAACCGAATTCACAAAGTT 4894  
Qy 264 ValSerLeuTyrLysLeuSerSerProGlu----- 273  
Db 4893 GTCAAAAGTTTCCGTAGATTAATGATGAGTGAAGTCTTCACATTAACGCAATTGTA 4834

Qy 274 -----AspAspProThrCysLysThrLysGluPheTPrAlaThrIleLeu 288  
Db 4833 GTTGTGTTGACCGGTGCTGACTGCTGCTCAAAAGTCAAGACTTAGATCCCTGTTGCG 4774  
Qy 289 AspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGluSerThr 308  
Db 4773 GATAGGCTCGGT-----GATTACGAGGCCGCCGCAAAATATGCGCGCTTCCTTA 4723  
Qy 309 ThrGly-----PheThrLeuTyrGlyMetLeuTyr--- 318  
Db 4722 TATGTTCTTAACTTTGTTGATCCGAAAGATTTCTTATTTGTTGTTGATCAACGG 4663  
Qy 319 -----LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeu 333  
Db 4662 CGGTACCTGACACCTAAACCTTTGGAGAAAGATGCGGAACACT----- 4618  
Qy 334 PheIleTyrGlyGlyProGlnValAlaIleAlaGlyAlaProValThrLeuTPrIlePhe 353  
Db 4617 TTCAAAATACGGG-----ATGTCAGTTGGCCGACGTAACCGACTGGAGATT 4573  
Qy 354 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAenGluGlnGlyTyr 373  
Db 4572 TACGATTCCTGTTTAACTGAGAGGTACATGATCTCTCAAGAAACTTTGATGATGATAC 4513  
Qy 374 TyrLeuGlySerVal-----AlaMetGlnAlaGluLysPheProSerGluPro 389  
Db 4512 GTAGATCAACGCTTCAATATGTCATGCTTTG-----GCACAGCA 4471  
Qy 390 AsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAenValHisPheAlaHisThrSer 409  
Db 4470 AATGATTTTGTGATCGACGAAACAGAGATGATTAACGTTCAATTCACAAATTCCTTA 4411  
Qy 410 IleLeuLeuSerPheLeuValArgAlaGly--LysProTyrAspLeuGlnIleTyrPro 428  
Db 4410 AAGTTTCTGGACCTTTTGATCTTAATGTTGTGAAATTTATGACGTCACGCTTCTCT 4351  
Qy 429 GlnGluArgHisSerIleArg 435  
Db 4350 GACTCAGATCATAGTATAGA 4330

RESULT 12  
US-09-280-116-172  
; Sequence 172, Application US/09280116A  
; Patent No. 6331427  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Pro tease Homologs  
; FILE REFERENCE: 5800-24, 035800/176965  
; CURRENT APPLICATION NUMBER: US/09/280,116A  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 268  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 172  
; LENGTH: 502  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: prolyl oligopeptidases  
US-09-280-116-172

Alignment Scores:  
Pred. No.: 1.89e-10 Length: 502  
Score: 167.00 Matches: 32  
Percent Similarity: 55.68% Conservative: 17  
Best Local Similarity: 36.36% Mismatches: 37  
Query Match: 6.73% Indels: 2  
Gaps: 1

US-10-070-464-5 (1-465) x US-09-280-116-172 (1-502)

Qy 346 AlaProValHisLeuTPrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHis 365  
Db 346 AlaProValHisLeuTPrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHis 365

Db 67 GCACCTATCACAGACTGAAATTGTATGCTTCTCTGAAAGATACCTTGGGATG 126  
QY 366 ProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlnLysPhe 365  
Db 127 CCATCTTAAGAGAAAGACCTTACGAGCCAGGTGTGTACATAATGTCATGCGCTTG 186  
QY 386 ProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPhe 405  
Db 187 AAGAGAGA-----AATATATTAAATTCATGAGACCTGACACAAAGTTTCATTTC 240  
QY 406 AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 425  
Db 241 CAACACTCAGCAGAAATTAATCAAGACCTAATAAAGCTGAGTGAATATATACATGACG 300  
QY 426 IleTyrProGlnGlnArgHisSer 433  
Db 301 GTCTACCCAGATGAAGGTCAATAC 324

RESULT 13  
US-09-221-017B-962  
; Sequence 962, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FASTSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/021, 017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1182  
; FILING DATE: 31-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1546  
; FILING DATE: 30-JAN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP2911  
; FILING DATE: 09-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU98/01023  
; FILING DATE: 10-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Monroy, Gladys H  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-813-5600  
; TELEFAX: 650-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 962:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 815 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: UNKNOWN  
; ORIGINAL SOURCE:  
; ORGANISM: PORYPHYROMONAS GINGIVALIS

; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1..815  
; US-09-221-017B-962

Alignment Scores:  
Pred. No.: 1,12e-07 Length: 815  
Score: 146.50 Matches: 67  
Percent Similarity: 37.33% Conservative: 42  
Best Local Similarity: 22.95% Mismatches: 114  
Query Match: 5.90% Indels: 69  
Gaps: 14

US-10-070-464-5 (1-465) x US-09-221-017B-962 (1-815)

QY 43 GlnTyrIleAlaArgAlaGlyTyrThrProGlnGlnLysTyrValAspSerIleLeuLeu 62  
Db 123 AAATTTCTGACGAATTTGAGTGTGAGTCCGAGCAAAATAATCTTGATGTAGTGAAGTGTG 182  
QY 63 AspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGlnLeu-----PheIle 80  
Db 183 AATCGTGCTCAAAACGAATGTAAAGTAAATGCTTATGACGCTGACCGGTAGATTGCTC 242  
QY 81 ProValGlnAspAspValMetGlnArgGlnArgLeuIleGlnSerValProAspSerVal 100  
Db 243 -----CGTACGCTTTTGTAAACCGATAACATTATGTA 278  
QY 101 ThrProLeuIleIleTyrGlnGlnThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
Db 279 GAGCCGCTTA-----CATCCCTGACA 299  
QY 121 HisValPheProGlnSerHisGlnGlnGlnIleGlnPheIlePheAlaSerGlnCysLys 140  
Db 300 TTCCTT---CCGGAAAGTAAACAAT-----CAGTTCAATTGGCGAGACCGCT---CGC 344  
QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGlnSerLysTyrLysArg 160  
Db 345 GACGAGTGAACCATCTCTATCTGTATGAT----- 374  
QY 161 SerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGlnIleAla 180  
Db 375 ACTACAGTCTGCTG-----ATCCGTGAC----- 398  
QY 181 IleThrSerGlyGlnThrProGlnValLeuGlyArgHisGlySerAsnIleGlnValAspGln 200  
Db 399 GTGACAAAGAGGAGTGGAGGTTTACAACTTTGCAAGC-----TTGCATCCC 446  
QY 201 ValArgArgLeuValTyrPheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyr 220  
Db 447 AAGGAAACAGCGCTTATTTCCAAAGTACCGAAGCCAGCCCTTCGAAAGCCCATTTTTC 506  
QY 221 ValValSerTyrValAsnProGlnGlnValThrArgLeuThrAspArgGlyTyrSerHis 240  
Db 507 TGTATTGAT---ATCAAGAGAGAAAGACAAAGATCTGACTCCGAGTGGGAAATGCAC 563  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
Db 564 CGCACTCACTATCTCTGATGAGTGTTCGCATTAATGATATTTTACAGTCACTACTGTC 623  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGlnAspArgProThrCysLysThr 280  
Db 624 CCGCGTAAGTTTACAGTGAACAAATATCGGCAAGGCTTCAC----- 665  
QY 281 LysGlnPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
Db 666 -----ACACTCTTGAGGCTAAGAACCCCGATACGGGCTATGCCAGTCCG 710  
QY 301 GlnIle-----PheSerPheGlnSerThrThrGlyPheThr---LeuTyrGlnMetLeu 317  
Db 711 GAGATCAGAACGGGATCATCATGATGCGGCGATGGGCGACAGCTTTATTATCAAGCTC 770  
QY 318 TyrLysProHisAspLeuGlnProGlnLysLysTyr 329

Db 771 ACGATGCCGCTTCATTTGATCCGGCAAGAAATAT 806

RESULT 14  
US-09-016-080-2  
Sequence 2, Application US/09016080  
Patent No. 6133012  
GENERAL INFORMATION:  
APPLICANT: Ishikawa, Kazuhiko  
APPLICANT: Matsui, Ikuo  
APPLICANT: Ishida, Hiroyasu  
APPLICANT: Kosugi, Yoshiyugu  
APPLICANT: Higuchi, Katsuniko  
TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING  
FILE REFERENCE: 07898/022001  
CURRENT APPLICATION NUMBER: US/09/016,080  
EARLIER APPLICATION NUMBER: JAPAN 18381/1997  
EARLIER FILING DATE: 1997-01-31  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1896  
TYPE: DNA  
ORGANISM: Pyrococcus horikoshii  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1896)  
US-09-016-080-2

Alignment Scores:  
Pred. No.: 3,5e-06 Length: 1896  
Percent: 139.00 Matches: 113  
Score Similarity: 30.58% Conservative: 65  
Best Local Similarity: 19.42% Mismatches: 172  
Query Match: 5.60% Indels: 232  
DB: Gaps: 28

US-10-070-464-5 (1-465) x US-09-016-080-2 (1-1896)

QY 7 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleAspVal 26  
Db 307 AAGTTCAATAATACGGGATAAAGAACCTCGCTTACCGAGATGGGAAAGATATAGCCGTG 366  
QY 27 ILe-----AspLysGluLeuIleGlnProPhe 35  
Db 367 GTTACCCCTATAGACGTTGAGAAAAGGAGATGATACGTTCACTATATAGGAAATA 426  
QY 36 GluIleLeuPheGluGlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLys 55  
Db 427 CCATTCTGTGTTTATAGAGTT-----GGCTGGATC--TAGGAAAA 465  
QY 56 TyrAlaTyrPheSerIleLeuLeuAsp-----ArgSerGlnThrArgLeu----- 69  
Db 466 AGAACCTGTCTACCTGTGTGACGTTGAGAGCGGAGAAAGAAAGACATTAACCTCAAG 525  
QY 70 -----GlnIleValIleSerProGluLeuPheIleProValGluAsp 84  
Db 526 AACCTAAATGTATGATCGATAGGTTCCACACGCGTATATACCTTACAGGCCCAAG 565  
QY 85 Asp-----ValMetGluArgGln----- 90  
Db 586 GATAGGGAGAAAGAACTCTGATATCCGATCTTACCTCTCGAGATTAAGAAAGTTAGG 645  
QY 91 -----ArgLeuIleGluSerValProAspSerValThrProLeu 103  
Db 646 AAGCTGACCCCAAGAGAGTGAAGATACCTGACCTCTCCCTTATGACGGAAGCTTC 705  
QY 104 IleIleTyrGluGluThrThrAsp-----IleTyrIleAsnIleHisAspIlePheHis 121  
Db 706 GTACTTAAGGCTAAACCTTTAGAAAGGGAGATCCCAACCAAGCCGCAC--ATCTACAC 762  
QY 122 ValPheProIleSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThr 141

Db 763 TACGATCCC-----AGACA 777  
QY 142 GlyPheArgHisIleuTyrIlySerIleLeuLysGluSerLysTyrLysArgSer 161  
Db 778 GGA-----GAACTTAAGAGCTCACAAAGATTTAGACAGAAAGCTTCAACCTCTTA 831  
QY 162 SerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIle 181  
Db 832 AAC-----TCCGATGTTCCAGGAAGTCAAGAGCGCCGACCTGTGTAC 873  
QY 182 ThrSerGlyGluTyrPheValLeuGlyArgHisGlySerAsnIleGlnValAspGluVal 201  
Db 874 AAGAGGGG--TGG----- 885  
QY 202 ArgArgLeuValTyrPheGluGluGlyThrLysAspSerProLeuGlnHisIleuTyrVal 221  
Db 886 -----ATCTACTATGTGCGACG--GATGGCCCT--AGGCAACCTCTTTAGG 930  
QY 222 ValSerTyrValAsnProGlyGluValThrArgLeuThr-----AspArgGlyTyrSer 239  
Db 931 GTCACTTA-----GATGGAAGATTTGAAGGGTATATAGTGAAGATTAGACGCTTGA 984  
QY 240 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259  
Db 985 -----AGCTTCGATATAGGGGATTACATAGCTTTCCAGC 1017  
QY 260 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspProThrCysLys 279  
Db 1018 GCTCAAGATGCTGTATACCCCACTGATCATATATACAGGATGGAAGAGAAAGAG 1077  
QY 280 ThrLysGluPhe-----TyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyr 297  
Db 1078 GTTACCGACTTTAAACAATG-----ATMAAGCGTTAC 1110  
QY 298 Thr-----ProProGluIlePheSerPheGluSerThrGlyPheThrLeuTyrGly 315  
Db 1111 ACCCTTCAAAACCTGAACACTTTAAGGTTAAAGCAAGTACGCGGGTTGAATATAGATGC 1170  
QY 316 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValIlePheIle 335  
Db 1171 TGGGTATATGAACCGGTGAACCTTCAGGAAAGAGAAAGATATCCAGCTATTTAGACATC 1230  
QY 336 TyrGlyGlyProGluValAlaIleAla----- 344  
Db 1231 CACGGTGGTCTAAACCGCTTACGGTTACGCTTTATATGACAGAGTTCCACGTTTAAAC 1290  
QY 344 ----- 344  
Db 1291 TCTAAAGGCTTCGTCGATATTTCTCAAACTTACAGAGAGCGATGGCTACGAGAGAG 1350  
QY 344 ----- 344  
Db 1351 TTCGGGATATTAAGGGAGACCTATGGGGAGAGGATTTACAGAGATTTATGAGAGTATGC 1410  
QY 344 ----- 344  
Db 1411 GATGAAGCATTAAGAGATTTGACTTCATAGATGGGAAAAGCTAGAGATCCGGCGGT 1470  
QY 345 -----GlyAlaProValThrLeuTyrPhePheTyrAspThrGlyTyrThrGluArgTyr 362  
Db 1471 TCCATATGGTGGCTTCATGACGAACTGATA-----GTCCGACATATCAACAGGTTCC 1521  
QY 362 ----- 362  
Db 1522 AAAGCGCGCTTAACCCAGAGATCAATTTCAATTGAGTAAAGCTTTCCGGGAACAAGGAT 1581  
QY 363 MetGlyHis-----ProAspGln-----AsnGluGlnGly 372  
Db 1582 ATAGTTATTAATCTTCTCCAGATCAAAATAGGAAAGAGATCCCTGAGCAACTTGAAGGT 1641  
QY 373 TyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeu 392



Db 1642 TATTGGGAAAAAGCCATTAAAGTAGCT-----CCCAAGTTGAAACTCCCTG 1692  
 Qy LeuLeuLeuH:sgLYPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeu 412  
 |||:|||||  
 Db 1693 CTTATATCCACTCTACCGAAGACTACAGGTGTGGCTTCCCGAGGCACTTGCACTCTTC 1752  
 Qy 413 SerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHis 432  
 |||:|||||  
 Db 1753 ATATCCCTTAAATACCTGGGAGAGAGATTGATTCGCAATATTCCTCCAGGAAAAATCAT 1812  
 Qy 433 SerIle 434  
 |||:|||||  
 Db 1813 GACCTA 1818

## RESULT 15

US-09-221-017B-253  
 ; Sequence 253, Application US/09221017B  
 ; Patent No. 6444799  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ross, Bruce C.  
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1120  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 755 PAGE MILL ROAD  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304-1018  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: Windows  
 ; SOFTWARE: FastSeq for Windows Version 2.0b  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/221,017B  
 ; FILING DATE: 23-DEC-1998  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PP1182  
 ; FILING DATE: 31-DEC-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PP1546  
 ; FILING DATE: 30-JAN-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PP2911  
 ; FILING DATE: 09-APR-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/AU98/01023  
 ; FILING DATE: 10-DEC-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Monroy, Gladys H  
 ; REGISTRATION NUMBER: 32,430  
 ; REFERENCE/DOCKET NUMBER: 27340-20021.00  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-813-5600  
 ; TELEFAX: 650-494-0792  
 ; TELEX: 706141  
 ; INFORMATION FOR SEQ ID NO: 253:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 543 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: circular  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYBRIDIZATION: NO  
 ; ANTI-SENSE: UNKNOWN  
 ; ORGANISM: PORPHYROMONAS GINGIVALIS  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 1...543  
 ; US-09-221-017B-253

## Alignment Scores:

Pred. No.: 2,61e-06 Length: 543  
 Score: 132.50 Matches: 30  
 Percent Similarity: 51.69% Conservative: 16  
 Best Local Similarity: 33.71% Mismatches: 40  
 Query Match: 5.34% Indels: 3  
 DB: 4 Gaps: 1

US-10-070-464-5 (1-465) x US-09-221-017B-253 (1-543)

Qy 340 GlnValAlaIleAlaGlyAlaProValThrLeuTyrPheTyrAspThrGlyTyrThr 359  
 |||:|||||  
 Db 283 AAAGTCGAGTAGAGCCGGCGCTGTCTATAGACTGGAATCGATATGAGATTATGACGT 342  
 Qy 360 GlnArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAla 379  
 |||:|||||  
 Db 343 GAGCGTTATTTTCGATGCCGCCACAGGAAAAATCCCGAAGATACGATGCTGCCAACTGCTC 402  
 Qy 380 MetGlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisgLYPheLeu 399  
 |||:|||||  
 Db 403 AAACGAGCCGGTGATCTG-----AAAGCAGCACTTATGCTATTCATGACGCGATC 453  
 Qy 400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419  
 |||:|||||  
 Db 454 GATCCGGTCGTCGTATGCGACGATTCATCTCTTTCTTGATGCTTGCGTGAAGGACGAC 513  
 Qy 420 LysProTyrAspLeuGlnIleTyrPro 428  
 |||:|||||  
 Db 514 ACCTATCCTGACTCTTACGTCATCCG 540

Search completed: October 16, 2003, 03:31:00  
 Job time : 100.851 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 16, 2003, 00:05:32 (search time 347.655 seconds  
(without alignments)  
3512.533 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTATPKVTFKMSSEIMDAE.....HLHYQENLGSRIALXVI 465

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1750203 seqs, 1333063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus\_model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10070464/runat\_15102003\_113555\_24924/app\_query.fasta\_1.2652  
-DB=Published Applications NA -QPMT=faeacp -SUFFIX=trpb -MIMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEARSTIE=500 -MINLEN=0  
-MAXLEN=2000000000 -USRR=US10070464 @CGN\_1.1.845 @runat\_15102003\_113555\_24924  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubna/PT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubna/PTTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubna/US10C\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2446.5	98.6	4523	US-09-976-674-8	Sequence 8, Appli

2	2422	97.6	2649	12	US-10-054-776-1	Sequence 1, Appli
3	2422	97.6	2649	12	US-10-170-789-39	Sequence 39, Appli
4	2422	97.6	2671	10	US-09-976-674-2	Sequence 2, Appli
5	2422	97.6	3143	12	US-10-170-789-37	Sequence 37, Appli
6	2418.5	97.4	4676	10	US-09-976-674-20	Sequence 20, Appli
7	2402	96.8	4829	10	US-09-976-674-12	Sequence 12, Appli
8	2149	86.6	4685	10	US-09-976-674-22	Sequence 22, Appli
9	1494.5	60.2	2617	10	US-09-976-674-4	Sequence 4, Appli
10	1494.5	60.2	4219	10	US-09-976-674-28	Sequence 28, Appli
11	1494.5	60.2	4302	10	US-09-976-674-24	Sequence 24, Appli
12	1445	58.2	4180	10	US-09-976-674-36	Sequence 36, Appli
13	1445	58.2	4263	10	US-09-976-674-34	Sequence 34, Appli
14	1386	55.8	4076	10	US-09-976-674-14	Sequence 14, Appli
15	1342.5	54.1	4076	10	US-09-976-674-32	Sequence 32, Appli
16	1340.5	54.1	4159	10	US-09-976-674-30	Sequence 30, Appli
17	1296	52.2	3262	13	US-10-098-841-83	Sequence 83, Appli
18	1293	52.1	4037	10	US-09-976-674-40	Sequence 40, Appli
19	1293	52.1	4120	10	US-09-976-674-38	Sequence 38, Appli
20	1276	51.4	2801	13	US-10-098-841-100	Sequence 100, App
21	471.5	19.0	2079	13	US-10-044-090-843	Sequence 843, App
22	471.5	19.0	2411	10	US-09-976-674-26	Sequence 26, Appli
23	401.5	16.2	561	11	US-09-764-891-877	Sequence 877, App
24	291	11.7	4797	11	US-09-764-891-7074	Sequence 7074, App
25	279	11.2	502	11	US-09-918-995-19585	Sequence 19585, A
26	267	10.8	281	9	US-09-867-550-987	Sequence 987, App
27	249.5	10.1	4835	10	US-09-917-800A-1570	Sequence 1570, Ap
28	249.5	10.1	4835	14	US-10-165-603-5	Sequence 5, Appli
29	247	10.0	2388	12	US-09-870-133-3	Sequence 3, Appli
30	247	10.0	2388	14	US-10-160-501-6	Sequence 6, Appli
31	247	10.0	2583	10	US-09-976-674-6	Sequence 6, Appli
32	247	10.0	3238	12	US-09-870-133-1	Sequence 1, Appli
33	247	10.0	3238	14	US-10-160-501-4	Sequence 4, Appli
34	247	10.0	4541	10	US-09-976-674-42	Sequence 42, Appli
35	238	9.6	3407	12	US-10-423-714-5	Sequence 5, Appli
36	238	9.6	3407	13	US-10-002-593-5	Sequence 5, Appli
37	238	9.6	3407	14	US-10-165-603-6	Sequence 6, Appli
38	216	8.7	2612	12	US-10-205-219-59	Sequence 59, Appli
39	214	8.6	2366	12	US-10-101-510-683	Sequence 683, App
40	214	8.6	2788	12	US-10-269-909-22	Sequence 22, Appli
41	214	8.6	2814	10	US-09-962-832-108	Sequence 108, App
42	214	8.6	2814	10	US-09-954-456-1148	Sequence 1148, Ap
43	214	8.6	2814	10	US-09-954-531-367	Sequence 367, App
44	214	8.6	2814	12	US-10-101-510-30	Sequence 30, Appli
45	214	8.6	2814	12	US-10-301-822-54	Sequence 54, Appli

#### ALIGNMENTS

RESULT 1  
US-09-976-674-8  
Sequence 8, Application US/09976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Akiyama, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 8  
LENGTH: 4523  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-976-674-8  
Alignment Scores: 1.1e-303 Length: 4523  
Pred. No.: 1

Score: 2446.50 Matches: 463  
Percent Similarity: 99.57% Conservative: 0  
Best Local Similarity: 99.57% Mismatches: 0  
Query Match: 98.57% Indels: 2  
DB: 10 Gaps: 1

US-10-070-464-5 (1-465) x US-09-976-674-8 (1-4523)

QY 1 ThrG1Thra1aenProLysVal1ThrPhelYmweSerGlu1IleMet1IleAsp1Aglu 20  
DB 1165 ACAGGTRACAGCAATCTTAAGTCATCTTTAAGTCACAGAAATTAAGATTGATGCTGAA 1224  
QY 21 GlyArlle1leasPVal1IleAspLysGluLeu1IleGlnProPheGlu1IlePheGlu 40  
DB 1225 GGAAGGATCATAGTCATAGATTAAGAACTAATTCACCTTTTGAGATTCTATTGAA 1284  
QY 41 GlyValGluTr1Ile1IleA1Agl1Agl1YTrPThrProGluGlySer1YTrA1ATPSe1Ile 60  
DB 1285 GGAAGTTGAATATATATGCGAGAGCTGAGTGAATCTCTAGGGAATATGCTGCTCATC 1344  
QY 61 LeuLeuAspA1RgSerGln1Thra1RgLeuGlu1IleValLeu1IleSerProGluLeuPhe1Ile 80  
DB 1345 CTACTAGATGCTCTCCAGACTCGCTACAGATGATGATCTCACTGAAATTTATTC 1404  
QY 81 ProValGluAspA1RgVal1MetGluA1RgGlnA1RgLeu1IleGluSerVal1ProAspSerVal1 100  
DB 1405 CCAAGTAGAATGATGATGATTAAGAAAGCAGAGACTCATTTGAGTGCCTGATTCTGTG 1464  
QY 101 ThrProLeu1Ile1IleYTrGluGlu1ThrThra1Ile1IleA1Ile1IleA1IlePhe 120  
DB 1465 ACCCCACTAATATATCTATGAAAGAAACACAGACATCTGGAATATTCATGACATCTTT 1524  
QY 121 HisValPheProGlnSer1IleSGluGluGlu1IleGluPhe1IlePhe1IleAsp1IleCysLeu 140  
DB 1525 CAGCTTTTCCCAAGATCAGAAAGAAATTAAGTTATTTTGCTCTGATGCTGAA 1584  
QY 141 ThrGlyPheA1Rg1IleLeu1YTrLys1IleThSer1IleLeuLysGluSerLys1YTrLysA1Rg 160  
DB 1585 ACAGGTTCCGCTATTTATACAAATTAATCATTTATTTAAAGAAACAAATATTAACGA 1644  
QY 161 SerSerGluGluLeuProA1AProSerA1RgPheLysCysPro1IleLysGluGlu1IleA1A 180  
DB 1645 TCCAGTGCTGAGCTGCTGCTCCAGAGATTTCAAGGTCCTATCAAAAGAGACATGCA 1704  
QY 181 I1eThSerGluGluTrP1Rg1IleVal1IleGluY1A1Rg1IleGlySerA1IleGlnVal1Asp1Ile 200  
DB 1705 ATTAACGATGCTGATGGAAGATCTTGCCCGCATGATTAATATCCAAAGTTGATGAA 1764  
QY 201 ValA1Rg1IleVal1YTrPheGluGly1YThLysA1RgSerProLeuGlu1IleHisLeu1YTr 220  
DB 1765 GTCAGAGAGCTGCTGATATTTTGAAGGCAACCAAGACTCCCTTTAGAGCATCACCTGTAC 1824  
QY 221 ValValSer1YTrVal1AenProGlu1YgluVal1Thra1RgLeu1Thra1RgGly1YTrSerHis 240  
DB 1825 GTAGTCAGTACGTAATCTCGAGAGGTGCAAGGCTGAGTACCGGCTACTCATCAT 1884  
QY 241 SerCysCys1IleSerGlnHisCysA1RgPhe1IleSerLys1YTrSerA1IleGlnLysA1A 260  
DB 1885 TCTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1944  
QY 261 ProHisCysVal1SerLeu1YTrLysLeuSerSerProGluAspA1RgPro1Thra1CysLeu1Thr 280  
DB 1945 CCAACACGTGCTGCTCTTACAGCTATCAAGTCTCAAGTATGACCAACCTGCAAAACA 2004  
QY 281 LysGluPhe1Rg1IleA1Rg1IleLeuAspSer1Agl1ProLeuProA1Rg1YTrThProPro 300  
DB 2005 AAGGAATTTTGGGCAACATTTTGATTCAGCAGGCTCTCTCTGATATATCTCTCTCA 2064  
QY 301 Glu1IlePheSerPheGluSer1Thr1GlyPhe1ThrLeu1YTrGlyMetLeu1YTrLysPro 320  
DB 2065 GAATATTTCTCTTTGAAAGTACTAGATTTACATTTGATGATGATGATGATGATGATGAT 2124  
QY 321 HisAspLeuGlnProGlyLys1YTrPro1Thra1LeuPhe1Ile1YTrGlyGlyProGln 340

DB 2125 CATGATCTACAGCTCGAAAGAAATATCCTAAGTCTGCTGTTCAATATGATGCTGCG---- 2180  
QY 341 ValAla1Ile1IleA1IleA1Rg1IleA1Rg1IleA1Rg1IleA1Rg1IleA1Rg1IleA1Rg1Ile 360  
DB 2181 GTTGCTATGCTGGGGCCAGATGATCTGTGATCTTCTATGATGATGATGATGATGATGATGAT 2240  
QY 361 Arg1YTrMetGlu1IleHisProAspGlnA1IleGluGlu1YTr1YTrLeuGluGluSerVal1A1Met 380  
DB 2241 CGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2300  
QY 381 GlnA1IleGlu1YTrPheProSerGluProAspA1RgLeuLeuLeuLeuLeuLeuLeuLeuLeu 400  
DB 2301 CAAGCAAGAAAGTTCCCTCTGAAACCAATCGTTTACGCTTACATGATGATGATGATGATGAT 2360  
QY 401 GluAspValHisPheA1IleHis1IleThSer1IleLeuLeuSerPheLeuValA1Rg1Agl1YTr 420  
DB 2361 GAGAAATGCTCAATTTTGACATACACAGTATATTAAGTATTTTAAAGAGGCTGGAAG 2420  
QY 421 Pro1YTrAspLeuGlu1Ile1YTrProGlnGluA1Rg1IleA1Rg1IleA1Rg1IleA1Rg1Ile 440  
DB 2421 CCATATGATTTTACAGATCTATCTCTCAAGAGACACAGCATTAAGATTTCTGAAATCGGA 2480  
QY 441 GluHis1YTrGluLeuHisLeuLeuHis1YTrLeuGlnGluAsnLeuGlySerA1IleA1A 460  
DB 2481 GAACATATGAACTGATCTTTTGACACTACCTTCAAGAAACCTTGATGATCAGTATTCCT 2540  
QY 461 AlaLeuLysVal1Ile 465  
DB 2541 GCTCTAAAGTGATA 2555

RESULT 2  
US-10-054-776-1  
: Sequence 1, Application US/10054776  
: Publication No. US20030165818A1  
: GENERAL INFORMATION:  
: APPLICANT: Mark Robert Edbrooke  
: APPLICANT: Alan Peter Lewis  
: TITLE OF INVENTION: NOVEL PROTEIN  
: FILE REFERENCE: QG1042US  
: CURRENT APPLICATION NUMBER: US/10/054, 776  
: CURRENT FILING DATE: 2002-01-23  
: NUMBER OF SEQ ID NOS: 2  
: SOFTWARE: PatentIn version 3.0  
: SEQ ID NO 1  
: LENGTH: 2649  
: TYPE: DNA  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: (1)..(2649)  
US-10-054-776-1

Alignment Scores:  
Pred. No.: 6,42e-301 Length: 2649  
Score: 2422.00 Matches: 465  
Percent Similarity: 82.30% Conservative: 0  
Best Local Similarity: 82.30% Mismatches: 0  
Query Match: 97.58% Indels: 100  
DB: 12 Gaps: 1

US-10-070-464-5 (1-465) x US-10-054-776-1 (1-2649)

QY 1 ThrG1Thra1aenProLysVal1ThrPhelYmweSerGlu1IleMet1IleAsp1Aglu 20  
DB 952 ACAGGTRACAGCAATCTTAAGTCATCTTTAAGTCACAGAAATTAAGATTGATGCTGAA 1011  
QY 21 GlyArlle1leasPVal1IleAspLysGluLeu1IleGlnProPheGlu1IlePheGlu 40  
DB 1012 GGAAGGATCATAGTCATAGATTAAGAACTAATTCACCTTTTGAGATTCTATTGAA 1071  
QY 41 GlyValGluTr1Ile1IleA1Agl1Agl1YTrPThrProGluGlySer1YTrA1ATPSe1Ile 60

Db 1072 GGAGTTGAATATATGTCAGAGCTGATGACTCTCGAGGAAATATGCTGTCATC 1131  
Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1132 CTACTGATGATGCTCCAGACTCGCTGCAGATAGTGTGATCTCACCTGAAATATTTATC 1191  
Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
Db 1192 CCAGTGAAGATGATGTTATGAAAGCAGAGCTACTTATAGTCAGTCCGTGATTCGTG 1251  
Qy 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
Db 1252 ACCCCACTATTTATCTATGAAAGAAACAAACAGACATCGAGTAAATATCCATGACATCTT 1311  
Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleAspGluCysGly 140  
Db 1312 CATGTTTTTCCCAAGTCAGAAAGAAATGAGTTATTTTGTGCTCTGATGACAA 1371  
Qy 141 ThrGlyPheArgHisLeuTyrIleTyrSerIleLeuIleGluSerIleTyrIleValArg 160  
Db 1372 ACAGGTTTCCGTCATTTATCAAAATTAACATCTATTTTAAAGAAAGCAATTAACCA 1431  
Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheIleCysProIleIleGluGluIleAla 180  
Db 1432 TCCAGTGGTGGCTGCTGCTCCAGATTTCAAGTGTCTATCAAGAGAGATAGCA 1491  
Qy 181 IleThrSerGlyGluIleTyrGluValLeuGluValArgHisGlySerAsnIleGlnValAspGlu 200  
Db 1492 ATTACAGTGGTGAATGGAGAGTTCTTGCCGCGAGTGAATCTATATTCGAAGTTGATGA 1551  
Qy 201 ValArgArgLeuValTyrPheGluGlyThrIleAspSerProLeuGluHisIleTyr 220  
Db 1552 GTCAGAAAGCTGGTATATTGTAAGGACCAAGACCTCCCTTAGGATCACTGTAC 1611  
Qy 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
Db 1612 GTAGTCAGTACCTAAATCTTGAGAGGAGCAAGGCTGACCTGACCTGACTACAT 1671  
Qy 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnIleValAsn 260  
Db 1672 TCTGCTGATCAGTACGACACTGTGCTCTTTATATAGTATATGTAACCAAGAAAT 1731  
Qy 261 ProHisCysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIleTyr 280  
Db 1732 CCACACTGTGTGCTCTTACAAAGCTATCAAGCTCGAAGATGACCCCACTTGCAAAACA 1791  
Qy 281 LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
Db 1792 AAGGAATTTGGGCCCAATTTGGATTCAGCAGGCTCTCTCTGACTATACTCTCCA 1851  
Qy 301 GluIlePheSerPheGluSerThrThrArgIlePheThrLeuTyrGlyMetLeuTyrIlePro 320  
Db 1852 GAAATTTTCTCTTTGAAAGTACTGATTTTCAATTTGTAAGGATGCTTACAAAGCCT 1911  
Qy 321 HisAspLeuGlnProGlyIleIleTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
Db 1912 CATGATCTACAGCTGGAAGAAATATCTACTGTGCTGTATATATGTTGTTCTCTGAG 1971  
Qy 340 ----- 340  
Db 1972 GTGAGTTGGTAATATGTTTAAAGAGTCAAGTATTCGCTTGAATACCTTAGCC 2031  
Qy 340 ----- 340  
Db 2032 TCTCTAGTTATGTGTTGTAGTATGACAAACAGGGGATCTGTGCACGAGGCTTAA 2091  
Qy 340 ----- 340  
Db 2092 TTGGAAGGCGCTTAAATATAATAATGGGTCAAAATAGAAATGACGATCAGGTGAAGA 2151  
Qy 340 ----- 340  
Db 2152 CTCCAATATCTAGCTTCTCGATATGATTTCAATTGACTTAGATCGTGTGGCATCGACGGC 2211

Qy 340 ----- 340  
Db 2212 TGCTCTAATGAGAGTAATCTCTCTCGATGCAATTAATGACAGGTGATATCTTCAGG 2271  
Qy 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
Db 2272 GTTGCTATTGCTGGGGCCCCAGTCACTGTGTGATCTTCTATATATACAGATACACGGA 2331  
Qy 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
Db 2332 CGTTATATGGGTCACTCCAGTCAAGATGAACAGGGCTATTAATAGATCTGTGGCATG 2391  
Qy 381 GluAlaGlyPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
Db 2392 CAAGCAGAAAGTTCCCTCTGAAACCAATCGTTACTGTCTTACATGGTTCTCGAT 2451  
Qy 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIle 420  
Db 2452 GAGATGTCATTTTGCACATACAGATATATTAATGATTTTATGAGGGCTGGAAG 2511  
Qy 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
Db 2512 CCATATGATTTACAGATCTATCTCTCAGAGAGACACAGCATAAAGATTCCTGAATCTGA 2571  
Qy 441 GluHisTyrGluLeuHisIleLeuLeuHisTyrIleGluGlnGluAsnLeuGlySerAlaGly 460  
Db 2572 GAACATTAATGAACCTGATCTTTTGACACTTCAAGTCAAGTGAATCACTGATTCCT 2631  
Qy 461 AlaLeuIleValIle 465  
Db 2632 GCTCTAAAGTGATA 2646  
RESULT 3  
US-10-170-789-39  
Sequence 39, Application US/10170789  
Publication No. US20030180930A1  
GENERAL INFORMATION:  
APPLICANT: Rachel E. Meyers  
APPLICANT: Olandt, Peter J.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
APPLICANT: Weich, Nadine  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
FILE REFERENCE: 10448-191001  
CURRENT APPLICATION NUMBER: US/10/170,789  
CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797,039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882,166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212,078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934,406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226,740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861,801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19

```

; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-170-789-39

Alignment Scores:
Pred. No.: 6,42e-301 Length: 2649
Score: 2422.00 Matches: 465
Percent Similarity: 82.30% Conservative: 0
Best Local Similarity: 82.30% Mismatches: 0
Query Match: 97.58% Indels: 100
DB: 12 Gaps: 1

US-10-070-464-5 (1-465) x US-10-170-789-39 (1-2649)
QY 1 ThGlyThrAlaAspPolyValThrPheLeuSerGluLeuMetIleAspAlaGlu 20
DB 952 ACAAGTACACCAATCTTAAAGTCATTTTAAAGTGCAGAAATATGATTGATGCTGAA 1011
QY 21 GlyArgIleIleAspValIleAspLeuGluLeuIleGlnProPheGluIleLeuPheGlu 40
DB 1012 GGAAGGATCATAGATGTCATAGATGAAGAACTAATTCACCTTTTGAGATTCTATTGAA 1071
QY 41 GlyValGluTyrIleAlaIleArgAlaGlyTyrThrProGluGlyLysTyrAlaIlePheSerIle 60
DB 1072 GGAGTTGAATATATTTGCGCAAGCTGATGATGATCTCTGAGGAAATATGCTGCTCATC 1131
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
DB 1132 CTACTAATATGCTCCACGACTCGCTCAATAGTGTGATCTCACTGAAATATATTTATC 1191
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
DB 1192 CCAAGTGAAGATGATGTTATGAAAGGACAGACTCATTTGAGTCAAGTCCGATTCCTG 1251
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrPheLeuIleHisAspIlePhe 120
DB 1252 ACCCCACTAATATATCTAAGAAACAACACATCTGATTAATATCCATGACATCTTT 1311
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140
DB 1312 CATGTTTTTCCCAAGTCAAGAAAGAAATGAGTTATTTTGGCTCTGATGCAAA 1371
```

```

QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160
DB 1372 ACAAGTTTCCGTCATTTATACAAATTTACATCTATTTTAAAGAAACCAATATTAACGA 1431
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluIleAla 180
DB 1432 TCCAGTGTGGGCTGCTGCTCCAGAGATTTCAAGTGTCTTCAAAAGAGAGATGCA 1491
QY 181 IleThrSerGlyGluTyrPheGluValIleGlyArgHisGlySerLeuIleGlnValAspGlu 200
DB 1492 ATTACCAAGTGTAAATGGAAGTTCTTGCGGCAATGATCTAATATCCAAAGTTGATGAA 1551
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220
DB 1552 GTCAGAGGCTGGTATATTTTGAAGCACCAAGACTCCCTTTAGACATCACTGTAC 1611
QY 221 ValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240
DB 1612 GTAGTCAGTTACGTAATCTCGAGAGGTACAAAGGCTGACTGACCGTGGCTACTCAT 1671
QY 241 SerCysGlyIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260
DB 1672 TCTTGCTGATCAGTCAAGTCAAGCTGTGACTCTTTAATAGTAATAGTAATCAAGAAAT 1731
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspProThrCysLysThr 280
DB 1732 CCACACTGTGTGCTCTTTCAAGCTATCAAGTCTGAAATATCAACCACTTGCAAAACA 1791
QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300
DB 1792 AAGGAATTTTGGGCAACCATTTTGATTCAGAGGTCTCTCTGCTGATACCTCTCA 1851
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320
DB 1852 GAAATTTTCTCTTTGAAATGACTACTGATTTTACATGATGATGAGAGCTCTCAACACCT 1911
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340
DB 1912 CATGATCTACAGCTGGAAGAAATATCTCATGCTGTGCTCATATATGATGATGCTCAG 1971
QY 340 ----- 340
DB 1972 GTGCAAGTGTGAATATCGGTTTAAAGAGATCAAGATTTCCGCTTGAATACCTTACCC 2031
QY 340 ----- 340
DB 2032 TCTCTAGTTATGTGTTGATGATGACACAGGGGATCTGTCAACGAGGCTTAA 2091
QY 340 ----- 340
DB 2092 TTTGAAGGCCCTTAAATATTAATGAGTCAATAGAAATTAACGATCAAGTGAAGAA 2151
QY 340 ----- 340
DB 2152 CTCCAATATCTAGCTTCTGATATGATTTCAATGATGATGATGATGATGATGATGATG 2211
QY 340 ----- 340
DB 2212 TGGTCTTAGGAGATACCTCTCTGATGAGCTTAATAGCAGAGGTCAAGATATCTTCAG 2271
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360
DB 2272 GTTGCTATGCTGGGCGCCAGACATCTGTGATCTTTATATATACAGATACACGAA 2331
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380
DB 2332 CGTTATATGGGTACCTCGACCAAGATGAACAGGGCTATTTACTTGAATCTGTGGCATG 2391
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAsp 400
DB 2392 CAAGCAAAAGTTCCCTCTGACCAATCGTTTACTGCTTTACATGATGTTCTTGAT 2451
```



Db 2399 CAGCAGAAAAGTCCCTGACCAATCGTTACTGCTTACATGGTTCCGAT 2458  
Qy 401 GIUASVNAHISHEALAHISHTSRISLEULEUSERHEULVALAAGLYLYS 420  
Db 2459 GAGAAATGTCATTTTGCATACACAGATATTAAGTTTAAAGAGGCTGAAAG 2518  
Qy 421 PROTYASPLEUGINLETYPROGLNUAGHISERILLEAGVALPROGLUSERGLY 440  
Db 2519 CCATATGATTTTACAGATCTTCTCCAGAGACACAGCTTAAGATTCTGAATCGGA 2578  
Qy 441 GIUHSITYGILUENHISLEULEHISITYLEUGLNUASNULEUGLYSERATGILEAIA 460  
Db 2579 GAACTATATGACATGCTTTTGGACTACTTCAAGAAAACCTGGATCAGGATTTGCT 2638  
Qy 461 ALALEULYSVALILE 465  
Db 2639 GCTCTAAAAGTGATA 2653

RESULT 5  
US-10-170-789-37  
Sequence 37, Application US/10170789  
Publication No. US20030180930A1  
GENERAL INFORMATION:  
APPLICANT: Rachel E. Meyers  
APPLICANT: Olande, Peter J.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
APPLICANT: Welch, Nadine  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF  
FILE REFERENCE: 10448-191001  
CURRENT APPLICATION NUMBER: US/10/170,789  
CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797,039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882,166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212,078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934,406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226,740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861,801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19  
PRIOR APPLICATION NUMBER: US 09/801,267  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07138  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,454  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/829,671  
PRIOR FILING DATE: 2001-04-10  
PRIOR APPLICATION NUMBER: PCT/US01/40483  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: US 60/197,508  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/961,721  
PRIOR FILING DATE: 2001-09-24

PRIOR APPLICATION NUMBER: PCT/US01/29904  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: US 60/235,023  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US 10/045,367  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: US 60/246,561  
PRIOR FILING DATE: 2000-11-07  
PRIOR APPLICATION NUMBER: US 09/801,275  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07074  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,420  
PRIOR FILING DATE: 2000-03-07  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 37  
LENGTH: 3143  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (229)...(2874)  
US-10-170-789-37

Alignment Scores:  
Pred. No.: 8,578-301 Length: 3143  
Score: 2422.00 Matches: 465  
Percent Similarity: 82.30% Conservative: 0  
Best Local Similarity: 82.30% Mismatches: 0  
Query Match: 97.58% Indels: 100  
Gaps: 1

US-10-070-464-5 (1-465) x US-10-170-789-37 (1-3143)

Qy 1 ThrGlyThrAlaAsnProIysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
Db 1180 ACAGGTACAGCAAAATCCTAAAGCATCTTTAAAGATGTCAGAAATATGATGCTGCA 1239  
Qy 21 GlyArgIleIleAspValIleAspLysGluLeuIleGluProPheGluIleLeuPheGlu 40  
Db 1240 GGAAGGATCATAGATGATCATAGTAAGAACTAATTAACCTTTAGATCTAATTTGAA 1299  
Qy 41 GlyValGluTyrIleIleAspAlaGlyTyrThrProGluGlyLysTyrAlaTSPSerIle 60  
Db 1300 GGAAGTGAATATTTCCAGAGCTGATGACCTCCAGAGAAATATGCTGCTCATC 1359  
Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1360 CTACTGATCGCTCCAGACTCCCTGCAGATGATGTTGATCTCACCTGAATATTTATC 1419  
Qy 81 ProValGluAspAspValMetGluArgGluIleGluSerValProAspSerVal 100  
Db 1420 CCAGTGAAGATGATTTATGAAAGCAGAGCTAATGAGTCAAGCTGATCTGTCG 1479  
Qy 101 ThrProLeuIleIleTyrGluGluIleThrAspIleTyrIleAsnIleHisAspIlePhe 120  
Db 1480 AGCCCACTAATATATCATGAAGAAACACAGACATCTGATTAATATCATGACATCTTT 1539  
Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
Db 1540 CAGTITTTCCCAAGATCAGAGAGAGAAATGAATTTATTTTCTCTGAAATCAAA 1599  
Qy 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
Db 1600 ACAGGTTTCGTCATTTATTAACAAATTAATCATCTATTTAAGAGAAAGCAATTAACGA 1659  
Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluIleAla 180  
Db 1660 TCCAGTGTGGGCTGCTCTCCAAAGTATTCAAAGTTCCTATCAAGAGGAGATGACA 1719  
Qy 181 IleThrSerGlyGluIleProIleValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200

Db 1720 ATTACAGTGTGATGGAAGTTCTTGCCGGCATGATCTAATATCCAAAGTTGATGAA 1779  
Qy 201 VALATGATGLeuVal1YrPheGluGlyThrLeuAspSerProLeuGluHisIleuTy 220  
Db 1780 GTTCAGAGGCTGGTATATATTTTGAAGGCACCAAGACCTCCCTTTAGCATCACTGTAC 1839  
Qy 221 VALATSerTyValAsnProGlyGluVal1ThrArgLeuThrAspArgGlyTySerHis 240  
Db 1840 GTATGATGATTAAGTAAATCTGAGAGGTGACAAAGCTGATCCCTGGCTACTCAT 1899  
Qy 241 SerCysHisIleSerGlnHisCysAspPhePheIleSerIleTySerAsnGlnIleAsn 260  
Db 1900 TCTTGTGCATCAGTACGACACTGTGACTCTTATATAAGTAAATAGTAAACCAAGAA 1959  
Qy 261 ProHisCysValSerLeuTyIleLeuSerSerProGluAspAspProThrCysIleThr 280  
Db 1960 CCACACTGTGTGCTCCCTTCAAGCATATCAAGTCTGAAATATACCACTTCACAAACA 2019  
Qy 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyThrProPro 300  
Db 2020 AAGGAATTTGGGCCACCATTTTGATTCAGAGCTCTTCTGACTATACCTCTCCA 2079  
Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyGlyMetLeuTyIlePro 320  
Db 2080 GAAATTTTCTCTTTGAAAGTACTACTGATTTACATTTATGAGATGCTCTCAAGCCT 2139  
Qy 321 HisAspLeuGlnProGlyIleValSerTyIleProThrValLeuPheIleTyGlyProGln 340  
Db 2140 CATGATCTACAGCTGGAAGAAATCTTACTGTGCTGTCAATATATGTTGTTCTCTCAG 2199  
Qy 340 ----- 340  
Db 2200 GTGCAGTTGGTGAATATCGCTTTAAAGAGTCAAGATTTCCGCTTGAATACCTAAGC 2259  
Qy 340 ----- 340  
Db 2260 TCTCTAGTTATGTGTGATGATGATGACAAAGGGAATCTGTACCGAGGGCTTAA 2319  
Qy 340 ----- 340  
Db 2320 TTTGAAGCGCCTTAAATATAAATGGCTCAATATGAAATGACATCAGTGGAGAA 2379  
Qy 340 ----- 340  
Db 2380 CTCCATATCTAGCTTCTGATATGATTTTATGACTTAGATGCTGTGGGATCCAGGC 2439  
Qy 340 ----- 340  
Db 2440 TGGTCTATGAGAGATACCTCTCCCTGATGACATTAATGACAGATCAGATATCTCAG 2499  
Qy 341 VALATATleAlaGlyAlaProVal1ThrLeuThrIlePheTyThrAspThrGlyTyThrGlu 360  
Db 2500 GTTGTCTATGTCTGGGCGCCAGTCACTGTGTGATCTTCTATATATCAAGATACAGGAA 2559  
Qy 361 ArgTyMetGlyHisProAspGlnAsnGluGlnGlyTyTyThrLeuGlySerValAlaMet 380  
Db 2560 CGTTATATGGGTACCTGACCAAGATGAACAGGGCTATTTACTTAAGATTTGTGGCCATG 2619  
Qy 381 GlnAlaGluIlePhePheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
Db 2620 CAAGCAGAAAAGTTCCCTCTGAACCAATCGTTTACTGCTTACATGGTTTCCGGAT 2679  
Qy 401 GluAsnValHisPheHisIleThrSerIleLeuLeuSerPheLeuValArgAlaGlyIle 420  
Db 2680 GAGAAATGTCATTTTACATACATACAGATATATTAAGTTTATATGAGGCTGGAAG 2739  
Qy 421 ProTyAspLeuGlnIleTyProGlnGluArgHisSerIleArgVal1ProGlySerGly 440  
Db 2740 CCATATGATTTACAGATCTATCTCTCAGAGAGACACAGATTAAGATTTCTGATCGGA 2799  
Qy 441 GluHisTyGluLeuHisIleLeuLeuHisTyThrLeuGlnGluAsnLeuGlySerArgIleAla 460  
Db 2800 GAACATTAAGAACTGCATCTTTTGGACATACCTTCAAGAAAACCTTGATCACTATTGCT 2859

Qy 461 AlaleuLeuValIle 465  
Db 2860 GCTCTAAAAGTGATA 2874  
RESULT 6  
US-09-976-674-20  
Sequence 20, Application US/09976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Aktinsanya, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPTV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
PRIORITY FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
PRIORITY FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 20  
LENGTH: 4676  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-976-674-20  
Alignment Scores:  
Pred. No.: 4,74e-300 Length: 4676  
Score: 2418.50 Matches: 463  
Percent Similarity: 90.10% Conservative: 1  
Best Local Similarity: 89.90% Mismatches: 1  
Query Match: 97.44% Indels: 51  
Gaps: 1  
US-10-070-464-5 (1-465) x US-09-976-674-20 (1-4676)  
Qy 1 ThrGlyThrAlaAsnProIleVal1ThrPheIleMetSerGluIleMetIleAspAlaGlu 20  
Db 1165 ACAGGTACAGCAAAATCTTAAGTCACTTAAAGATGTCAGAAATATGATTTGCTGAA 1224  
Qy 21 GlyArgIleIleAspValIleAspIleGluLeuIleGluProPheGluIleLeuPheGlu 40  
Db 1225 GGAAGATCATATGATTCATATGATTAAGAACTAATTCACCTTTGAGATTTCAATTTGAA 1284  
Qy 41 GlyValGluTyIleAlaArgAlaGlyTyThrProGluGlyTySerTyAlaThrPheIle 60  
Db 1285 GAGGTGAATATATTTCCAGAGCTGATGAGCTCCAGAGGAAATATGCTGTGCATC 1344  
Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1345 CTACTAGATGCTCCAGAGCTGCCCTAAGATGTGATCTCACTGAATATTATTC 1404  
Qy 81 ProValGluAspAspValMetGluArgGluIleGluSerVal1ProAspSerVal 100  
Db 1405 CCAGTGAAGATGATGTATGAAAGCAGAGCTATGAGTCACTGCTGATTCGTG 1464  
Qy 101 ThrProLeuIleIleTyGluGluThrThrAspIleTyIleAsnIleHisAspIlePhe 120  
Db 1465 ACCGCCATATATCTATATGAAGAAACACAGACATCGATTAATATTCATGACATCTTT 1524  
Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIle 140  
Db 1525 CATGTTTTTCCCAAGTACAGAAAGGAAATGAGTTATTTTCCCTCAATACCAAA 1584  
Qy 141 ThrGlyPheArgHisIleuTyIleValIleThrSerIleLeuIleGluSerTyIleValArg 160  
Db 1585 ACAGGTTTCCGATTTATTAACAAATATACATCTATTTTAAAGGAAAGCAAAATTAACGA 1644  
Qy 161 SerSerGlyIleLeuProAlaProSerAspPheIleCysProIleIleGluGluIleAla 180  
Db 1645 TCCAGTGGTGGGCTGCTGCTCCAAAGTATTTCAAGTGTCTATCAAAAGAGGATAGCA 1704



```

QY 181 ILeThrSerGlyGluTrpGluValIleuGlyArgHisGlySerAsnIleGlnValAspGlu 200
    |||||
DB 1705 ATTACCAAGTGGTGAATGGAGAGTTCTTGCGCGGCAATGATATCCAAATGATGAA 1764
QY 201 ValArgArgLeuValIlePheGluGlyThrLeuAspSerProLeuGlnHisIleuTy 220
    |||||
DB 1765 GTCAGAAAGCTGGTATATTTTGAAGGCAACCAAGACTCCCTTTAGAGCATCACTGTAC 1824
QY 221 ValValSerTyValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240
    |||||
DB 1825 GTAGTCAGTACCTGAAATCTGAGAGAGTACAAAGGCTGACCTGGCTACTCATCAT 1884
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerTyTyrSerAsnGlnLysAsn 260
    |||||
DB 1885 TCTTGCTGCATCACTGCACACTGTGACTTTTATAGTAAGTATAGTAAACCAAGAAAT 1944
QY 261 ProHisCysValSerLeuTyThrLeuSerSerProGluAspAspProThrCysValSer 280
    |||||
DB 1945 CCACACTGTGTGCTCCCTTACAAAGCTATCAAGCTCTGAAAGATGACCCAACTTGCAAAACA 2004
QY 281 LysGluPheThrPheValThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300
    |||||
DB 2005 AAGCAATTTTGGCGCACCATTTGGATTGACAGAGTCTCTCTGACTAATACCTCTCCA 2064
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyThrIleuLeuTyThrLysPro 320
    |||||
DB 2065 GAATTTTCTCTTTTGAATACCTACCTGATTTTACATTTGATGGAGCTCTCAAGACCT 2124
QY 321 HisAspLeuGlnProGlyLysLysTyProThrValLeuPheIleTyThrGlyGlyPro 339
    |||||
DB 2125 CATGATCTACAGCTGGAAAGAAATATCTCATGTGCTTCATATGTGTGTGCG-GGT 2183
QY 339 ----- 339
DB 2184 CAAATGAAATTCAGATCAGTCAGTGAAGGACTCCAAATATCTAGCTTCCATATGATTC 2243
QY 339 ----- 339
DB 2244 ATGACTTATAGTGTGTGGGATCCACGCTGGTCTCATGAGAGATACCTCTCCCTGATG 2303
QY 340 -----GlnValAlaIleAlaGlyAlaProValThrLeu 350
    |||||
DB 2304 GCATTAATGACAGAGTCAGATATCTTCAGGGTGTCTATTTCTGGGGCCCAAGTCACCTGT 2363
QY 351 TrpIlePheTyArgPheThrGlyTyThrGluArgTyThrMetGlyHisProAspGlnAsnGlu 370
    |||||
DB 2364 TGGATCTTCTATGATACAGATACAGAAAGTTATATGGGTCACTGACCAAGATGAA 2423
QY 371 GlnGlyTyThrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsn 390
    |||||
DB 2424 CAGGGCTATTAAGTATGATCTGTGGCATGCAAGCAAAAGTTCCCTCTGAAACCAAT 2483
QY 391 ArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIle 410
    |||||
DB 2484 CGTTTACTGCTTACTACGTGTTCCTGAGAGAGATTCATTTTGGACATACACAGATATA 2543
QY 411 LeuLeuSerPheLeuValArgAlaGlyLysProTyArgPheGlnIleTyThrProGlnGlu 430
    |||||
DB 2544 TTACTAGTCTTTTATGAGAGGCTGGAACCAATATGATTTACATGATCTATCTCTGAGAG 2603
QY 431 ArgHisSerIleArgValProGluSerGlyGlnHisTyThrGluLeuHisLeuLeuHisTy 450
    |||||
DB 2604 AGACACAGCATTAAGTTCCTGATTCGGAGAACATATGACATGATCTTTTGGACTAC 2663
QY 451 LeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 465
    |||||
DB 2664 CTTCAAGAAACCTTGGATCACTATGCTGCTTAAAGTGAATA 2708

```

RESULT 7  
 US-09-976-674-12  
 ; Sequence 12, Application US/0976674  
 ; Patent No. US20020115843A1

```

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akimaneva, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 12
; LENGTH: 4829
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-976-674-12

Alignment Scores:
Pred. No.: 6,7e-298 Length: 4829
Score: 2402.00 Matches: 465
Percent Similarity: 82.01% Conservative: 0
Best Local Similarity: 82.01% Mismatches: 0
Query Match: 96.78% Indels: 102
DB: 10 Gaps: 1

US-10-070-464-5 (1-465) x US-09-976-674-12 (1-4829)

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
    |||||
DB 1165 ACAGTACAGCAAAATCTTAAGTCACTTTAAGATGTCAGAAATATGATGTGCTGAA 1224
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProGluIleLeuPheGlu 40
    |||||
DB 1225 GGAAGATCATATGATGTCATAGTAAAGAACTTAATCACTTTGAGATTCATTTGAA 1284
QY 41 GlyValGluTyThrIleAlaArgAlaGlyTyThrProGluGlyLysTyAlaThrSerIle 60
    |||||
DB 1285 GGAGTGAATATATTTCCAGAGCTGATGATGACTCCGAGGAAATATGCTTGCTCATC 1344
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
    |||||
DB 1345 CTACTGATGCTGCCAGACTCCCTACAGATGATGTGATCTCACCTGAATATTATTC 1404
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
    |||||
DB 1405 CCAGTGAAGATGATGTATGGAAGGCAAGACTGATGAGTCAAGTGCCTGATTCGTG 1464
QY 101 ThrProLeuIleIleTyThrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePhe 120
    |||||
DB 1465 AGCCACTAATATATCTATGAGAAACACAGACATGTGATTAATATTCATGACATCTTT 1524
QY 121 HisValPheProGlnSerHisGlyGluGluIleGlnPheIlePheAlaSerGluCysLys 140
    |||||
DB 1525 CATGTTTTTCCCAAGTACAGAAAGAAATGAGTTATTTTCTCTGATGCCAA 1584
QY 141 ThrGlyPheArgHisLeuTyThrIleThrSerIleLeuLysGluSerLysTyThrLysArg 160
    |||||
DB 1585 ACAGTTCCTCGTATTTATACAAATATTAATCTATTTTAAAGAAACCAATATTAACGA 1644
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180
    |||||
DB 1645 TCCAGTGTGGGCTGCTGCTCCAAAGTATTCAAAGTCTCTATCAAGAGAGATAGCA 1704
QY 181 ILeThrSerGlyGluTrpGluValIleuGlyArgHisGlySerAsnIleGlnValAspGlu 200
    |||||
DB 1705 ATTACCAAGTGGTGAATGGAGAGTTCTTGCGCGGCAATGATATCCAAATGATGAA 1764
QY 201 ValArgArgLeuValIlePheGluGlyThrLeuAspSerProLeuGlnHisIleuTy 220
    |||||
DB 1765 GTCAGAAAGCTGGTATATTTTGAAGGCAACCAAGACTCCCTTTAGAGCATCACTGTAC 1824

```

QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 1825 GTAGTCACTACGTAATTCCTGGAGAGTGAACAAGCTGACTGACCTGCTACTCAT 1684  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerIleSerAsnGlnLysAsn 260  
 DB 1885 TCTTGCTGCATCAGTCACTGACCTGCTCTTTAATAAGTATAGTAACTCAAGAGAAAT 1944  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 1945 CCACACGTGTGTCCTTACAAAGCTATCAAGTCTGAAGATGACCAACTTCGCAAAACA 2004  
 QY 281 LysGluPheIlePheIleThrIleLeuAspSerIleGlyProLeuProAspTyrThrProPro 300  
 DB 2005 AAGCAATTTTGGGCCACCATTTTGGATTGACAGAGTCTCTTCTGACTATCTCTCCCA 2064  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 2065 GAAATTTTCTTTTGAAGTACTACTGATTTACATTTATGAGATGCTCTCAAGCCT 2124  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--ProG 340  
 DB 2125 CATGATCTACAGCTGGAAAGAAATATCTACTGTGCTGTTCATATATGATGATCTCTC 2184  
 QY 340 1n----- 340  
 DB 2185 AGGTGAGTGTGTAATATCGGTTTAAAGAGTCAAGTATTTCCGTTGAATACCTAG 2244  
 QY 340 ----- 340  
 DB 2245 CCTCTAGGTATGTGTGTATGATATAGCAACAGGGAGTCTGTACCGAGGCTTA 2304  
 QY 340 ----- 340  
 DB 2305 AATTGAAAGCGCCTTAAATATATAAATGGCTCAAAATGAAATTGACGATCAGGTGAAAG 2364  
 QY 340 ----- 340  
 DB 2365 GACTCCAAATATCAGCTTCTCGATATGATTTTATGACTTATGATCGTGGGATCCACG 2424  
 QY 340 ----- 340  
 DB 2425 GCTGCTCATGAGAGATACCTCTCCCTGATGCAATTAAAGAGTCCAGATATCTTCA 2484  
 QY 341 --ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrG 360  
 DB 2485 GGGTTCCTATTTGCTGGGGCCCGACGTCCTGTGGATCTTCTATGATACAGGATACACG 2544  
 QY 360 LysArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLysGlySerValAla 380  
 DB 2545 AACGTTATATGGGTCAACCTGACCAAGAAATGAACAGGGCTATTAAGATCTGTGGCA 2604  
 QY 380 eTGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 400  
 DB 2605 TCCAAAGCAAAAAAGTCCCTCTGAAACCAAAATGTTACTGTGCTTACATGATGTTCTCTG 2664  
 QY 400 ArgGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyAla 420  
 DB 2665 ATGAGAGATTCATTTTGCACATACCAATATATATCTGAGTTTAAAGAGGCTGGA 2724  
 QY 420 LysProTyrAspLeuGlnIleTyrProGlnGlnLysArgHisSerIleArgValProGluSerG 440  
 DB 2725 AGCCATATATGATTTACAGATCTATCTCCAGAGAGACACAGCATTAAGATCTCTGAATCGG 2784  
 QY 440 LysGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnLysSerArgIleAla 460  
 DB 2785 GGAACATATATGAACGATCTTTTGGCACTACCTTCAAGAAAACTTGGATCACGATATG 2844  
 QY 460 1aAlaLeuLysValIle 465  
 DB 2845 CTGCTCTAATAAGTGATA 2861  
 RESULT 8

US-09-976-674-22  
 ; Sequence 22, Application US/09976674  
 ; Patent No. US20020115843A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Qi, Steve  
 ; APPLICANT: Akinsanya, Karen  
 ; APPLICANT: Riviere, Pierre  
 ; APPLICANT: Junien, Jean-Louis  
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
 ; FILE REFERENCE: 70669  
 ; CURRENT APPLICATION NUMBER: US/09/976,674  
 ; PRIORITY FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: US 60/240,117  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 22  
 ; LENGTH: 4685  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-976-674-22  
 Alignment Scores:  
 Pred. No.: 2,6e-265 Length: 4685  
 Score: 2149.00 Matches: 428  
 Percent Similarity: 83.65% Conservative: 7  
 Best Local Similarity: 82.31% Mismatches: 28  
 Query Match: 86.58% Indels: 58  
 DB: 10 Gaps: 6  
 US-10-070-464-5 (1-465) x US-09-976-674-22 (1-4685)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 1165 ACAGGTACAGCAAAATCTTAAGTCACTTTAAGATGTCAGAAATATGATGATGCTGAA 1224  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 1225 GGAAGGATCATAGATGTCATATGATTAAGAACTTAATTCACCTTTTATGATTTATTTGAA 1284  
 QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 DB 1285 GGGATGAAATATATATGTCAGAGCTGATGATGATCTCCGAGGAAATATGCTTGCTCATC 1344  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 1345 CTACTAGATCGCTCCAGACTGCGCTACAGATATGTTGATCTCACCTGAATATTTATTC 1404  
 QY 81 ProValGluAspAspValMetGluLysArgLeuIleGluSerValProAspSerVal 100  
 DB 1405 CCAAGTGAAGATGATGTTATGGAAGGACAGACATTCATTAAGTCACTGCTGATTTCTGTG 1464  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 1465 AGCCACATATATATCTATGAAAGAAACACAGACATCTGATTAATATTCATGACATCTTT 1524  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 DB 1525 CATGTTTTTCCCAAGTACAGAGAGGAAATGAGTTATTTTGGCTCTGATGCAAA 1584  
 QY 141 ThrGlyPheArgHisLeuTyrIleThrSerIleLeuLysGluSerIleTyrLysArg 160  
 DB 1585 ACAGGTTTCCGTCATTTATACAAATTTATCATCTATTTTAAAGAAAGCAAAATATTAACGA 1644  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 DB 1645 TCCAGTGGTGGGCTCCTGCTCCAGATGATTTCAAGTGTCTCTATCAAGAGAGATAGCA 1704  
 QY 181 IleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 DB 1705 ATTACCAAGTGGTGAATGGAAGTTCTTGCGCGCATGATCTTAATATCAAGATGATGAA 1764  
 QY 201 ValArgGluLeuValTyrPheGluGluTyrHisAspSerProLeuGluHisIleLeuTyr 220

```

Db      1765 GTCAGAGGCTGGTATATTGTAAGGACCAAGAGCTCCCTTAAAGCATCAGCTGAC 1824
QY      221 ValValSerTyrValAsnProGlyGlyValThrArgLeuThrAspArgGlyTyrSerHis 240
Db      1825 GTAAGTAGTACCTAAATCTGGAGAGGTACAAAGGCTGACCGTGGCTACTACAT 1884
QY      241 SerCysCysIleSerGlnHisCysAspPheIleSerTyrSerAsnGlnIysAsn 260
Db      1885 TCTTGTGATCATGATGACGACCTGACTTCTTATAGTAGATAGTAAACGAGAAAT 1944
QY      261 ProHisCysValSerLeuTyrIleuSerSerProGlnAspAspProThrCysIysThr 280
Db      1945 CCACACTGTGTGTCCTCTTACAGAGTATCAAGCTGAGATGATGACCACTTCAAAACA 2004
QY      281 LysGlnPheThrPheIleThrIleuAspSerAla----- 291
Db      2005 AAGGAATTTTGGCCACCAATTTGGATTCAGTCTTCAAGTGACGAGTGGATGATATTCGGT 2064
QY      292 -----GlyProLeuProAspTyrThrProPro 300
Db      2065 TTAAGAGTCAAGTATTTCCGCTTGATATCCCTAGCTCTCTA--GGTATATGCGTTGTA 2123
QY      301 GluIle-----PheSerPheGluSerThrArgPhe 311
Db      2124 GTGATGAGACAACAGGGATCTGTCAACGAGGCTTAAATTTGAAGCGCTTAAATAT 2183
QY      312 ThrLeuTyrGlyMetLeuTyrIleProHisAspLeuGlnProGlyIleIysIleTyrProThr 331
Db      2184 AAAATG--GGTCAATAA--GAAATTGACATCAGGTGAGAGGAGCTCCCAATATCTAGCT 2237
QY      332 ValLeuPhe-----IleTyrGlyGly 338
Db      2238 TCTCGATATGATTCTATGACTTAGATTCGTGTGGGATCCAGGCTGTGCTTATGAGAG 2297
QY      339 Pro-----GlnValAlaIleAlaGly 345
Db      2298 TACCTCTCCCTGATGCAATTATGACAGAGTCAGATATCTTCAGGGTGTCTATGTCTGG 2357
QY      346 AlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHis 365
Db      2358 GCCCGAGTCACTGTGGATCTTCTATGATACAGATACCGGAAAGCTTATATGGGTAC 2417
QY      366 ProAspGlnAsnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyIlePhe 385
Db      2418 CCGTACACAGATTAACAGGCTATTTACTTAGATCTGTGGCCATGACAGAGAAAGTTCC 2477
QY      386 ProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPhe 405
Db      2478 CCTCTGAACCAATGTTTACTGCTCTTACATGATGTTTCTGTGATAGAAATGTCATTTT 2537
QY      406 AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIleProTyrAspLeuGln 425
Db      2538 GACATACACAGATATTTACTGAGTTTTTTAGTAGAGGCTGGAAAGCCATATGATTTACAG 2597
QY      426 IleTyrProGlnGluArgHisSerIleArgValProGluSerGlyIleHisTyrGluLeu 445
Db      2598 ATCTATCTCAGAGAGACACAGCATTAAGATTCTGAATCGGAGAAACATTTATGAATCG 2657
QY      446 HisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleuIysValIle 465
Db      2658 CATCTTTTGCATCACTTCAAGAAACCTTGGATCAGATATGCTCTCTTAAAGATGATA 2717

```

RESULT 9  
US-09-976-674-4  
Sequence 4, Application US/0976674  
Patent No. US20020115843A1

GENERAL INFORMATION:  
APPLICANT: OJ, Steve  
APPLICANT: Akinsanya, Karen  
APPLICANT: Riviere, Pierrie  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV

```

/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976,674
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 4
/ LENGTH: 2617
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-976-674-4

Alignment Scores:
Pred. No.: 2,24e-181
Score: 1494.50
Percent Similarity: 62.958
Best Local Similarity: 50.548
Query Match: 60.21%
DB: 10 Gaps: 2

US-10-070-464-5 (1-465) x US-09-976-674-4 (1-2617)

QY      1 ThrGlyThrAlaAsnProIleValThrPheIleMetSerGluIleMetIleAspAlaGlu 20
Db      932 ACAGGACAGCAAGAAATCCCAAGATTGCGTGAATGCTGAGATTCCAGACTGCACAGCAG 991
QY      21 GlyArgIleIleAspValIleAspIleAspIleGlnProPheGluIleLeuPheGlu 40
Db      992 GGCAGAGATGCTTCGACCCAGCAAGAGAGGTGTGTGACCCCTTCAAGCTGCTGTTCCCG 1051
QY      41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyIleTyrAlaTyrSerIle 60
Db      1052 AAGGTGAGATACATCCCGAGGCTGGGTGACCCGGAGTGGCAAAATCCCTGGGCTG 1111
QY      61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleIleSerProGluLeuPheIle 80
Db      1112 TTCCTGACACCGGCCCCAGCATGCTCCAGCTGCTCTCTCCCGCCGCTGTTCAATC 1171
QY      81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
Db      1172 CCGAGCACAGAAATGAGAGAGAGCGCTAGCCTCTCCAGAGAGTGTCCCAAGAAATGTC 1231
QY      101 ThrProLeuIleIleTyrGlnGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120
Db      1232 CAGCCTATGTGTGTACAGAGAGGTACCAACGCTGTGATCAATGTTTATGATCATCTTC 1291
QY      121 HisValPheProGlnSerHis--GluGluGluIleGluPheIlePheAlaSerGluCys 139
Db      1292 TATCCCTCCCAATCAAGAGGAGAGAGAGCTCTCTCTCCGCGCCCAATGAATGC 1351
QY      140 LysThrGluPheArgHisIleuTyrIleTyrSerIleLeuIleGluSerIleTyrIle 159
Db      1352 AAGACCGGCTTCTCCCATTTGTAACAAAGTACCGGCGTTTAAATCCCAAGGCTAGCAT 1411
QY      160 ArgSerSerGlyIleuProAlaProSerAspPheIleCysProIleIleGluGluIle 179
Db      1412 TGAAGTAGACCTTCAGCCCGGGAAGATGAATTTAAAGTCCCATTAAGAGAGATTT 1471
QY      180 AlaIleThrSerGlyIleuThrPheIleuValIleuGlyArgHisGlySerAsnIleGlnValAsp 199
Db      1472 GCTTGACCAAGGAGGAGGAGGTGTTGGAGAGGACCGGCTCCAAAGATCTGGGTCAT 1531
QY      200 GluValArgArgLeuValTyrPheGluGluTyrIleAspSerProLeuGlnHisIleu 219
Db      1532 GAGAGACCAAGCTGTGATCTTCCAGGACCAAGAGACCCCGCTGAGACCAACCTTC 1591
QY      220 TyrValIleSerTyrValAsnProGlyGluValThrArgLeuThrAspArgIleTyrSer 239
Db      1592 TACGTGTGATGAGAGGCGGCGGAGATCGATACGCTTCAACACGCGCGGCTTCTCC 1651
QY      240 HisSerCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnIys 259

```

Db	1652	CATAGCTGCTCATGATGAGCCAGAACTTCACAGATGTTGTCAGACCACTAAGACAGCGTGCAC	1711
Qy	260	AenPrOHisLeCysValSerLeuTyrTyrLeuSerSerProGluAerPaPProThrcYLeys	279
Db	1712	ACGCGCGCCCTGTGGTGCACGTCTTACACAGCTGAGCGCGCCCGCCGACGACGACCCCTGCACAG	1711
Qy	280	ThrlYsGluPheTPrAlaThrIleLeuAerSerAlaGlyProLeuProAspTyrThrPro	299
Db	1772	CAGCCCCGCTTGTGGGCTTACATGATGAGGAGGACGACGTGCCCCCGGATTAATGTTCT	1831
Qy	300	ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrlys	319
Db	1832	CCAGAGATCTTCATTTCCACACGCGGCTCGGATGTGCGGCTCTTACGGCATGATCTACAG	1891
Qy	320	ProH1saPheLeuGlnProGlyTyrLeuTyrTyrProThrValIleuPheIleTyrGlyPro	339
Db	1892	CCCCACGCGCTTGACGCGAGGAGAGAACCCACCGTCTCTTGTATATGAGGCGCC	1951
Qy	340	Gln-----	340
Db	1952	CAGGTGCAAGCTGTGATATACCTCTTCAAAAGGCATCAAGTACTTGCGGCTCAACACACTG	2011
Qy	340	-----	340
Db	2012	GCCTCCCTGGGCTACGCGGTGTGTGATGACGCGAGGCGCTCTGTCAAGCGAGGCTT	2071
Qy	340	-----	340
Db	2072	CGGTTCGAAAGGGGCGCCCTGAAAAACAAATTGGCCAGGCTGAGATCGAGACCAAGTGGAG	2131
Qy	340	-----	340
Db	2112	GAGCCTGCAATTGCTGGCCGAGACGATGATGCTTTCATCGACTGAGCGGAGTTGCCATCAT	2191
Qy	340	-----	340
Db	2192	GAGCTGGCTTACGGGGGCTTCTCTCGCTCATGGGGCTAATCCACAGCCCGAGTGTTC	2251
Qy	341	---ValAlaIleAlaGlyAlaProValThrLeuTPrIlePheTyrAerThrGlyTyrThr	359
Db	2252	AAGGTGGCATTCGGGGGTGCCCGGCTGACCGTGTGAGTGGGCTTACGACACAGGTTACACT	2311
Qy	360	GluATgTyrMetGlyHisPheAspGlnAsnGlnGlyTyrTyrLeuGlySerValAla	379
Db	2312	GAGCGCTTCAATGAAAGCTCCCTGAGAAACACAGACCGGCTATGAGGGCGGTTCCGTGGCC	2371
Qy	380	MetGlnAlaGluTyrPheProSerGlnProAsnAtrGleuLeuLeuHisGlyPheLeu	399
Db	2372	CTGCACGTGGAGAACGTCCCAATGAGCCCAACGGCTTCTTCTTCCACAGGCTTCTCG	2431
Qy	400	AerGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly	419
Db	2432	GACGAAAAACGTGACATTTTCCACACAAACTTCCGTCCTCCCAACTGATCGAGCAGGG	2491
Qy	420	LysProTyrAerLeuGlnIleTyrProGlnIuAtrHisSerIleAtrValProGluSer	439
Db	2492	AAACCTTACACAGCTTCCAGATCTACCCCAACGAGACACAGATATTCCCTCCCGAGTGC	2551
Qy	440	GlyGluHisTyrGluLeuHisLeuLeuHisTyrTyrLeuGlnGlnIuAsnLeu	455
Db	2552	GGCGAGCATATGAAGTACAGCTTCTGTGACTTTTCTACAGAAATACCTC	2599
RESULT 10			
US-09-976-674-28			
; Sequence 28, Application US/09976674			
; Patent No. US20020115843A1			
; GENERAL INFORMATION:			
; APPLICANT: Qi, Steve			
; APPLICANT: Akinsanya, Karen			
; APPLICANT: Riviere, Pierre			
; APPLICANT: Junien, Jean-Louis			
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV			
; FILE REFERENCE: 70669			

```

: CURRENT APPLICATION NUMBER: US/09/976,674
: CURRENT FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: US 60/240,117
: PRIOR FILING DATE: 2000-10-12
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 28
: LENGTH: 4219
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-976-674-28

Alignment Scores:
Pred. No.: 5,01e-181 Length: 4219
Score: 1494.50 Matches: 281
Percent Similarity: 62.95% Conservative: 69
Best Local Similarity: 50.54% Mismatches: 105
Query Match: 60.21% Indels: 101
DB: 10 Gaps: 2

US-10-070-464-5 (1-465) X US-09-976-674-28 (1-4219)

QY 1 ThrGlyThrAlaAspProValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
DB 1288 ACAGGACACAGAAATCCCAAGATTGCTTGAACATCGGTGAATTCACACTACAGCCAG 134
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleuPheGlu 40
DB 1348 GGCAAGATCGTCTGCACCCAGAGAAAGAGCTGTGACGCCCTTACGCTCGCTGTCCCG 140
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaIlePseIle 60
DB 1408 AAGGTGAGTACATGCCAGGCGCGGTGAGACCCGGATGGCAATACGCTGGGCGCATG 146
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleLysSerProGluLeuPheIle 80
DB 1468 TTCTCGGACCGGCCCCACAGTGGCTTCAGCTCTCTCCCTCCCGCCGCTGTTCATC 152
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
DB 1528 CCGAGCACAGAAATGAGAGACAGGCGGTACGCTCTGCCAAGCTGTCCCAAGAAATGC 158
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120
DB 1588 CAGCGGTATGGTGTACAGAGAGGTCCACCAACGTGTGATCAATGTTCAATGACATCTTC 164
QY 121 HisValPheProGlnSerHis--GluGluGluIleGluPheIlePheAlaSerGluCys 139
DB 1648 TATCCCTTCCCCCATCTCAGAGGAGAGGACGAGCTCTCTTCTCCGGCCAAATGATGC 170
QY 140 LysThrGlyPheAspGlnIleLeuTyrIleValIlePheSerIleLeuLysGluSerLysTyrLys 159
DB 1708 AAGACCGGCTTCTCCCATTTGTACAAATGACCGCGCTTTAAATCCAGAGGCTACCAT 176
QY 160 ArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIle 179
DB 1768 TGGAGTAGACCCCTCAGCCCGCGGGAAGATAAATTAAATGATCCCATTAAGAGAAAGATT 182
QY 180 AlaIleThrSerGlyGluTyrGluIleValLeuGluValArgHisGlySerAsnIleGlnValAsp 199
DB 1828 GCTCTGACCAAGCGGTGAATGAGAGATTGTCGAGGACCGGCTCCAAATCTCGGTCAAT 188
QY 200 GluValArgAspGluValTyrPheGluGlyTyrIleAspSerProLeuGluHisIleLeu 219
DB 1888 GAGAGACCAAGCTGGTGTATCTTCACAGGACCAAGAGACAGCGCTGAGACACCACTC 194
QY 220 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 239
DB 1948 TAGCTGTCAAGCTATGAGCGCGCGGCGAGATGCTAGAGCTCACACACCGCGGCTTCTCC 200
QY 240 HisSerCysValIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259
DB 2008 CATGTGCTTCCATGACCCAGAACTTGCATGTGTTCTGACGCCCTACAGAGGCTGAGC 206

```



Oy	260	asnpRhoHisCySValSerLeuTyrLysLeuSerSerProGluAaPaPProthThCybLys	275
Db	2068	ACGCCGCCCTGCGTGCACGCTCACAGCTGAGCGGCCCGGACGACGACACCCCTGCACAAAG	2127
Oy	280	ThrlYsGluPheTrpAlaThrIleLeuAAsPseAlaGluProLeuProAAsPTrpThrPro	299
Db	2128	CAGCCCCCGCTTCTGGGCTAGCATGATGAGAGCAGCAGCTGCCCCCGGATTAATGTTCTT	2187
Oy	300	ProGluIlePheSerPheGluSerThrTrpGlyPheThrLeuTyrGlyMetLeuTyrLys	319
Db	2188	CCAGAGATCTTCATTCATTCACACCGCGCTCGATGCGGCGCTCTACAGGATGATCTACAAAG	2247
Oy	320	ProHISaPLeuGlnProGlyLysLysTyrProThrValIlePheIleTyrGlyGlyPro	339
Db	2248	CCCCACGCCCTTGCAACCCAGGAGAGAACCCCAACCGTCCCTTGTATATGAGAGGCCCC	2307
Oy	340	Gln-----	340
Db	2308	CAGGTGCAGCTGTGTGAATTAATCTCTTCAAAAGGATCAAGTACTGGGCTCAACACTG	2367
Oy	340	-----	340
Db	2368	GCCTCCCTGGGGTACGCGGTGGTTGTATGACGCGACAGGGGCTCTGTACAGAGAGGCTT	2427
Oy	340	-----	340
Db	2428	CGTTTCGAAGGGGCCCTTGAAAAACCAATGGGCGCAGGTGAGATCGAGACCAAGACAGTGGAG	2487
Oy	340	-----	340
Db	2488	GAGCTGCAGTTCTGTGTGCCGAGAAAGTATGGCTTCATCGACTTGAGCCGAGTTGCCATCCAT	2547
Oy	340	-----	340
Db	2548	GGCTGTCTTACGGGGGGCTTCTCTCGCTCATGGGGGCTAATCCAAAGCCCAAGTGTTC	2607
Oy	341	---ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAAsPThrGlyTyrThr	359
Db	2608	AAGGTGGCCATCGCGGTCGCCCGGTCAACGCTTGATGGAGGCTTCAGACACAGGGTACACT	2667
Oy	360	GluAAsPTrpMetGlyHisProAAsPLeuAAsGlnGlnGlyTyrTyrLeuGlySerValAla	379
Db	2668	GAGGCGTACATGAGACGTCCTCTGAGAAACAACAGCAGCGCTATGAGCGCGGTTCCGTGGCC	2727
Oy	380	MetGlnAlaGluLysPheProSerSerGluProAAsPArgLeuLeuLeuHisGlyPheLeu	399
Db	2728	CTGCACGTCGAGAAAGCTCCCAATATAGCCCAACCGCTTCTTATCTCTCCAGCGCTTCTGTG	2787
Oy	400	AspGluAAsValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly	419
Db	2788	GACGAAACCGTGCACTTTTCCACACAACTTCCGTCTCCCAATGATCCGAGCAGAGGG	2847
Oy	420	LysProTyrAAsPLeuGlnIleTyrProGlnLysArgHisSerIleArgValProGluSer	439
Db	2848	AAACCTTAACCAAGCTCCACATCTACCCCCAACAGAGACACACTATTTCGTGCCCGAGTCTG	2907
Oy	440	GlyGluHisTyrGlyLeuLeuHisLeuLeuHisTyrLeuGlnGlnAAsLeuLeu	455
Db	2908	GGCGAGCACTATGAAGTACAGCTTGCTGCACTTTCTTACAGGAATACCTC	2955

RESULT 12

US-09-976-674-36

; Sequence 36, Application US/09976674

; Patent No. US20020115843A1

GENERAL INFORMATION:

; APPLICANT: Q1, Steve

; APPLICANT: Akinsanya, Karen

; APPLICANT: Riviere, Pierre

; APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976, 674

CURRENT FILING DATE: 2001-10-12

```

; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 4180
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-36

Alignment Scores:
Pred. No.: 1,18e-174 Length: 4180
Score: 1445.00 Matches: 274
Best Similarity: 61.33% Conservative: 67
Best Local Similarity: 49.28% Mismatches: 101
Query Match: 58.22% Indels: 114
De: 10 Gaps: 3

US-10-070-464-5 (1-465) x US-09-976-674-36 (1-4180)

QY 1 ThrGlyThrAlaAsnProLyseValThrPheLyseMetSerGluIleMetIleAspAlaGlu 20
Db 1288 ACAGGACACAGAAAGATCCCAATATGCGCTTGAAGAACTGCTGATTCACAGCTGACAGCCAG 1347
QY 21 GlyArgIleIleAspValIleAspLyseGluLeuIleGlnProPheGluIleLeuPheGlu 40
Db 1348 GGCAGAAATGCTGCGACCCAGAGAGAGAGCTGGTGCGAGCCCTTCAGCTGCTGTTCCCG 1407
QY 61 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLeuTyrAlaTyrSerIle 60
Db 1408 AAGGTGGAGTGCATCGCCAGGCGCGGGTGGACCCGGATGGCAAAATACGCTGGGGCCATG 1467
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
Db 1468 TTCCTGGACCGGCCCCAGCAGCTGGCTCCAGTCTGCTCTCCGCCCGGCTGTTATTC 1527
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
Db 1528 CCGAGCACAGAAATGAGAGCAGCAGCGGCTACCTCTGCCAAGACTGTGCCACAGAAATTC 1587
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleValIleHisAspIlePhe 120
Db 1588 CAGCGGTATGGTGTCAGAGAGAGTCCACCACTGGATCAATGTTCATGACATCTTC 1647
QY 121 HisValPheProGlnSerHis---GluGluGluIleGluPheIlePheAlaSerGluGly 139
Db 1648 TATCCCTTCCCCCATCGACGAGGAGAGACAGACTGCTGCTTCTCCCGGCAATGAAGTGC 1707
QY 140 LysThrGlyPheAsnGHisLeuTyrLeuIleHisSerIleLeuLysGluSerIleTyrLys 159
Db 1708 AAGACCGGCTTCTGCGCATTTTGTACAAAGTCAACCGCGCTTTTAAATCCAGGGCTACAT 1767
QY 160 ArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLeuGluGluIle 179
Db 1768 TGGAGTGAAGCCTTCAGGCCCGGGGAGAGATGAATTTAAGTGCCCCCATTAAGGAAGAGTT 1827
QY 180 AlaIleThrSerGlyGluTyrGluValLeuGluGlyArgHisGlySerAsnIleGlnValAsp 199
Db 1828 GCTCTGACACACGGGTGAATGGAGAGTATTGGCCAGGACCGCTCC----- 1872
QY 200 GluValArgArgLeuValTyrPheGluGlyThrIlyAspSerProLeuGluHisIleLeu 219
Db 1873 -----AAGGGACCAAGAGACAGCGGCTGGAGACACACCTC 1908
QY 220 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 239
Db 1969 TACGGTGTCAGCTATGAGGGCGGGGAGATGCTACAGCCACACAGCGCGGCTTCTCC 1968
QY 240 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerIlyTyrSerAsnGlnLys 259
Db 1969 CATGCTGCTCTCCATGAGGACAGAAATGTCAGACTGTTCTTCAGCCACTACAGACGCTGAGC 2028
QY 260 AsnProHisCysValSerLeuTyrValLeuSerSerProGluAsnAsnProThrCysLys 279

```



```

Db      2029 ACGCCGCCCTGCGTGCAGCTTAACTGAGCGGCCCGCAGCAGCAACCCCTGCAACAG 2088
Qy      280 ThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 299
Db      2089 CAGCCCCGCTTCTGGGCTACGATGATGAGGAGCGAGCTGCCCGCGGATTAATGTTCTT 2148
Qy      300 ProGluIlePheSerPheGluSerThrGlyPheThrLeuTyrGlyMetLeuTyrLys 319
Db      2149 CCAGAGATCTTCCATTTCCACAGCGCTGGATGTGGGCTCTACGCGCATGATCTAACAG 2208
Qy      320 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 339
Db      2209 CCCACGCGCTTGCAGCCAGGAGAAAGAACCCACCGCTCTTGTATATGAGGCGCCC 2268
Qy      340 Gln----- 340
Db      2269 CAGGTGCAGCTGTGTAATACTCTTCAAAAGCATCAAGTACTTGGCGGCTCAACACACTG 2328
Qy      340 ----- 340
Db      2329 GCCTCCCTGGGCTACGCGCGTGTGTGATGATGACGGCAGGGGCTTCTGTCAAGGAGGCTT 2388
Qy      340 ----- 340
Db      2389 CGGTTTGAAAGGCGCCTGTAACAAATGGCCAGGTGAGATCGAGACACAGTGGAG 2448
Qy      340 ----- 340
Db      2449 GGCCTGCACTTGTGGCCGAGAAATGCTGCTTATGACTGACCTGAGCCGAGTGGCCATCAT 2508
Qy      340 ----- 340
Db      2509 GGCTGGCTCTACGAGGGGCTTCTCTGCTCATGAGGGCTAATCCAAAGCCCGAGTGTTC 2568
Qy      341 ---ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 359
Db      2569 AAGGTGGCCATCGCGGGGTCCCGGTCACCGCTGTGATGGCTTACACACAGGGTACACT 2628
Qy      360 GluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlnIleSerValAla 379
Db      2629 GAGCGCTACATGAGCGCTCCCTTGAGAACACACAGCAGCGGCTATGAGCGGGTTCGAGGCC 2688
Qy      380 MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399
Db      2689 CTGCAAGTGGAGAGCTGCCCAATGAGCCCAACCGCTTCTTATCTCCACGCGCTTCCG 2748
Qy      400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419
Db      2749 GACGAAACGTGCACATTTCACACACAACTTCCTGCTCCCACTGATCCGAGCAGGAG 2808
Qy      420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
Db      2809 AAACCTTACCAAGCTCCAGATCTACCCCAACGAGACACAGATGCTGCGCCCGAGTGG 2868
Qy      440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 455
Db      2869 GCGCAGCACTATGAGTGCAGTGTGCTGCACTTCTTACAGAAATACTTC 2916

RESULT 14
US-09-976-674-14
; Sequence 14, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: OJ, Steve
; APPLICANT: Akimsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Jumeau, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
```

```

; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 4309
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-14

Alignment Scores:
Pred. No.: 5,02e-167 Length: 4309
Score: 1386.00 Matches: 270
Percent Similarity: 72.58% Conservative: 0
Best Local Similarity: 72.58% Mismatches: 0
Query Match: 55.84 Indels: 102
DB: 10 Gaps: 1

US-10-070-464-5 (1-465) x US-09-976-674-14 (1-4309)

Qy      186 ILeGlnValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 215
Db      1230 ATCCAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1289
Qy      216 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 235
Db      1290 GAGCATCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1349
Qy      236 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPheIleSerLysTyr 255
Db      1350 CGTGGCTACTGACATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1409
Qy      256 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerProGluAspAsp 275
Db      1410 AGTACCAAGAAATCAACAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1469
Qy      276 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 295
Db      1470 CCAACTTGCAGAAACAAAGAAATTTGGGCGACCATTTGGATTACACAGGCTCCTTCTCT 1529
Qy      296 AspTyrThrProProGluIlePheSerPheGluSerThrGlyPheThrLeuTyrGly 315
Db      1530 GACTATATCTCTCCACAAATTTCTCTTTGAAAGTACATGATTTACATGATTTAGTGGG 1589
Qy      316 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 335
Db      1590 ATGCTTCAAGGCTTATATATCTACAGCTTGGAAAGAAATATCTACTGCTGTTCAATA 1649
Qy      336 TyrGlyGly--ProGln----- 340
Db      1650 TATGGGTCTCTCTACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1709
Qy      340 ----- 340
Db      1710 GCTTGAATACCTAGCTCTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1769
Qy      340 ----- 340
Db      1770 GTCAAGGAGGCTTAATTGAAGGCGCTTAATAATATAAAATGGTCAAAATGAAATTG 1829
Qy      340 ----- 340
Db      1830 ACGATCAGGTGAGAGATCCCAATATCTAGCTTTCGATATGATTTGATTGATTGATTGATTGAT 1889
Qy      340 ----- 340
Db      1890 GTTGGGATCCACGCGTGTCTTATGAGAGATACCTCCCTGATGCGATTATGACAGA 1949
Qy      341 ---ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyr 355
Db      1950 GGTGATATCTTCCAGGGTGTCTATGCTGGGGCCCGCAGTCACTGTGATCTTCTATG 2009
Qy      355 sPTThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrL 375
Db      2010 ATACAGATACACGAGAGCTTATATGAGGTCAACCTGACCAAGATGAAACAGGCTATTACT 2069
```



QY 375 euGlYserValAlaMeGlnAlaGluYAspHeProSeGluProAsnArgLeuLeuL 395  
DB 2070 TAGAGTCTGTGGCATGCAAGCAAGAAAGTTCCCTCTGACCAATCGTTACTGCTC 2129  
QY 395 euHISglYpHeLeuAspGluAsnValHISpHeAlaHISrSerIleLeuLeuSerPheL 415  
DB 2130 TACATGGTTTCTCGATGAGAAATGTCATTTTGCATACACATATATATCTAGATTTT 2189  
QY 415 euValArgAlaGluYAspProGluYAspLeuGlnIleTyProGlnGluArgHISerIleA 435  
DB 2190 TACTGAGGCTGGAAGGCCATATGATTTACAGATCTATCTCAGAGAGACACAGCATAA 2249  
QY 435 rGValProGluSerGluYLeuHISrYgluLeuHISleuLeuHISrYLeuGlnGluAsnL 455  
DB 2250 GAGTCTCTGAAATCGGAGAACATTTATGACTGCATCTTTTGCACCTACCTTCAGAAAC 2309  
QY 455 euGlySerArgIleAlaLeuLeuYValIle 465  
DB 2310 TTGGATCAGCTATTTGCTGCTTAAAGTGATA 2341

RESULT 15  
US-09-976-674-32  
Sequence 32, Application US/09976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Akinsanya, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 32  
LENGTH: 4076  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-976-674-32

Alignment Scores:  
Pred. No.: 1,856-161 Length: 4076  
Score: 1342.50 Matches: 259  
Percent Similarity: 64.83% Conservative: 71  
Best Local Similarity: 50.88% Mismatches: 125  
Query Match: 54.09% Indels: 54  
DB: 10 Gaps: 4

US-10-070-464-5 (1-465) x US-09-976-674-32 (1-4076)

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMeSerGluIleMetIleAspAlaGlu 20  
DB 1288 ACAGGAGCAAGAAATCCCAAGATTGCTTGAACCTGCTGAGTTCACAGACTGCACGCG 1347  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProHISerGluIleLeuPheGlu 40  
DB 1348 GCGAAGTCTCTCGACCCAGCAAGAAAGAGCTGTGACGCCCTTCAGCTGCTGTTCCCG 1407  
QY 41 GlyValGluTyrlleAlaArgAlaGlyTyrlleProGluGlyLysTyrlleAlaTrpSerIle 60  
DB 1408 AAGGTGAGATACATCCGCAAGGCGCGGATGACCCGGAGTGCAGAAATACGCTGGGCGATG 1467  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValleuIleSerProGluLeuPheIle 80  
DB 1468 TTCCTGACCGGCGCCAGCAGAGTGGCTTCAGCTGCTCTCCCTCCCGGCGCTGTTTCATC 1527  
QY 81 ProValGluAspArgValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1528 CCGAGACAGAGAAATAGAGAGCAGCGGCTGAGCTCTGCCAGAGCTGTCCCAAGAAATGTC 1587

QY 101 ThrProLeuIleIleTyrlleGluGluThrThrAspIleTrpIleAsnIleHISerIlePhe 120  
DB 1588 CAGCCCTATGTGTGTACAGAGAGTACCAAGCTGTGATCATATGTTATGATCATCTTC 1647  
QY 121 HISValPheProGlnSerHIS---GluGluGluIleGluPheIlePheAlaSerGluCys 139  
DB 1648 TATCCCTCCCGCCCAATCAGAGGAGAGAGAGAGAGCTGCTTCTCCGCGCAATGAAATGC 1707  
QY 140 ILeThGlyPheArgHISleuTyrlleHISerIleLeuIleGluSerLysTyrlle 159  
DB 1708 AAGACCGGCTTCGCAATTGTACAAAGTCACCGCGTTTAAATCCAGGGCTACGAT 1767  
QY 160 ArgSerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluIle 179  
DB 1768 TGGAGTACAGCCCTTACAGCCCGGAGAAATGAAATTAAGTCCCATTAAGAAAGATG 1827  
QY 180 AlaIleThSerGlyLysGluValLeuGlyArgHISglYSerAsnIleGlnValAsp 199  
DB 1828 GCTCTGACAGCGGTGAATGGAGGTTTGGCGAGCCAGCGCTCAAGATCTGGGTCAAT 1887  
QY 200 GluValArgArgLeuValTyrllePheGluGluTyrlleLysAspSerProLeuGlnHISleu 219  
DB 1888 GAGAGACCAAGCTGTGTACTTCCAGGACACCAAGACACGCGCTGGAGACACACTC 1947  
QY 220 TyrlleValSerTyrlleAsnProGlyGluValThrArgLeuThrAspArgIleTyrlleSer 239  
DB 1948 TACGTGTACAGTATAGGCGCGCGCGAGATGTGACGCTCACACAGCCCGCTTCTCC 2007  
QY 240 HISerCysGlyLeSerGlnHISerLysAspPheHISerLysTyrlleSerArgGlnLys 259  
DB 2008 CATAGCTGTCTTCCATGACCAAGCTTCGACATGTTGTCACGACCATACAGCACGCTGAGC 2067  
QY 260 AsnProHISerCysValSerLeuTyrlleLysSerProGluAspArgProThrCysLys 279  
DB 2068 AGCGCCGCTGTGACGTCTACAGCTGAGGCGCCGAGACACACCCCTGACACAG 2127  
QY 280 ThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrllePro 299  
DB 2128 CAGCCCGCTTCTGGGCTGACATGATGACAGCGACGACGCTCCCGCATTTATGTTCT 2187  
QY 300 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrlleMetLeuTyrlle 319  
DB 2188 CCGAGATCTTCATTCACACAGCGCTCGATGTGCGCTTACGCGATGATCTTCAAG 2247  
QY 320 ProHISerLeuGluProGlyLysLysTyrlleProThrValLeuPheIleTyrlleGlyPro 339  
DB 2248 CCCACGCTTGTGACGCAAGGAGAAAGACCCACCGCTCTTGTATATGAGGCCCC 2307  
QY 340 GlnValAlaIleAlaGlyAlaPro-----ValThrLeuTrpIlePhe 353  
DB 2308 CAGGTGACAGCTGTGAATAAATCTTCAAGGATCAAGTACTTGGGCTCAACACAGT 2367  
QY 354 TyrllePheThrGlyTyrlleThr----- 359  
DB 2368 GCCTCCCTGGGTACCGCTGTGTGATTGACGAGGAGGCTCTGTGACGAGGCTT 2427  
QY 360 -----GluArgTyrlleMetGlyHISerProAspGlnAsnGln 371  
DB 2428 CGGTTGAAGGGGCTTGAATAAACAATGGGCAAGTGGAGTGAAGACAGGTGAG 2487  
QY 371 ----- 371  
DB 2488 GCGCTGCAATGCTGTGCGAGAAATAGGCTTCATGACGACGACGAGGATTCATCAT 2547  
QY 372 -----GlyTyrlleGluLysSerValAlaMet-GlnAlaGluLysPheProSe 387  
DB 2548 GCGTGTCTTACAGGGGCTTCTCTCGCTACAGGGGCTTATCCAAAGGCCCAAGGTGTC 2607  
QY 387 rGluProAsnArgLeuLeuLeuHISglYpHeLeuAspGluAsnValHISpHeAlaHIS 407  
DB 2608 AAGGCCCAACCGCTTGTCTTATCTTCCAGCGCTTCTGAGCAAAAGTGCATTTTCCA 2667



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:24:12 (Search time 2793.69 Seconds)

(without alignments)  
4045.398 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKVTRKMSSEIMIDAE.....HLHYIGENGSRIALXVI 465

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh  
-Q=/cgr2\_1/USPTO.spool/US10070464/runac\_15102003\_113553\_24829/app\_query.fasta\_1.2652  
-DB=EST -QFMT=fastap -SUFFIX=xst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsm62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10070464.@CCN 1.1.8056@runac\_15102003\_113553\_24829 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

EST: \*  
1: em\_estba: \*  
2: em\_estbm: \*  
3: em\_estin: \*  
4: em\_estm: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: em\_gse\_hum: \*  
18: em\_gse\_inv: \*  
19: em\_gse\_pln: \*  
20: em\_gse\_vrt: \*  
21: em\_gse\_fun: \*  
22: em\_gse\_mam: \*  
23: em\_gse\_mus: \*  
24: em\_gse\_pro: \*  
25: em\_gse\_rod: \*  
26: em\_gse\_png: \*  
27: em\_gse\_vrl: \*  
28: gb\_gse1: \*

29: gb\_gse2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2326	93.7	3143	11 AK016546	AK016546 Mus muscu
2	2234	90.0	5517	11 AK029788	AK029788 Mus muscu
3	1488.5	60.0	3327	11 AK050021	AK050021 Mus muscu
4	1488.5	60.0	3376	11 AK050023	AK050023 Mus muscu
5	1478.5	59.6	3457	11 AK078301	AK078301 Mus muscu
6	1397	56.3	910	13 BX372276	BX372276 BX372276
7	1265.5	51.0	853	12 B1223892	B1223892 602941035
8	1202	48.4	993	10 BG259714	BG259714 602380072
9	1157.5	46.6	1021	12 BG974587	BG974587 602844820
10	1127.5	45.4	1021	14 CB233750	CB233750 ACBNCOURT
11	1117.5	45.0	1042	12 BM557438	BM557438 ACBNCOURT
12	1082	43.6	1041	12 BI084090	BI084090 602869453
13	1063	42.8	762	14 CD352521	CD352521 UI-M-G10-
14	1055	42.5	753	9 AL040398	AL040398 DKF2P344A
15	1040	41.9	831	12 BG974324	BG974324 602844075
16	1034.5	41.7	957	13 BG675006	BG675006 ACBNCOURT
17	1028	41.4	956	14 CB153590	CB153590 K-EST0211
18	1015	40.9	910	13 BG675260	BG675260 ACBNCOURT
19	1015	40.9	968	13 BG671635	BG671635 ACBNCOURT
20	1011	40.7	897	13 BU916138	BU916138 ACBNCOURT
21	981.5	39.5	1085	9 AL582206	AL582206
22	972	38.2	575	10 BG086878	BG086878 H3131F11-
23	946	38.1	896	10 BF784154	BF784154 602108015
24	912	36.7	921	13 BU147109	BU147109 ACBNCOURT
25	880.5	35.5	1082	13 BU239476	BU239476 60332338
26	864	34.8	682	10 BM473819	BM473819 BB473819
27	857	34.5	1090	12 BM10838	BM10838 ACBNCOURT
28	850	34.2	893	10 BF694553	BF694553 602081767
29	849.5	33.2	845	13 BU216915	BU216915 603109373
30	831	33.5	469	9 AA465309	AA465309 aa24b08.x
31	811.5	32.7	943	13 BO895618	BO895618 ACBNCOURT
32	802.5	32.3	888	13 BU145896	BU145896 ACBNCOURT
33	797	32.1	658	10 BE388695	BE388695 601286345
34	792	31.9	467	14 CA541559	CA541559 CO607E01-
35	760	30.6	789	10 BG709118	BG709118 602675382
36	746	30.1	785	12 BI148072	BI148072 602912419
37	719	29.0	674	10 BE888665	BE888665 601513061
38	712	28.7	746	14 BY751026	BY751026 BY751026
39	709.5	28.6	902	13 BQ642814	BQ642814 ACBNCOURT
40	690.5	27.8	855	12 BI084885	BI084885 602869453
41	682.5	27.5	727	12 BI855677	BI855677 603383206
42	678	27.3	735	13 BU447277	BU447277 603766685
43	673.5	27.1	970	10 BG403555	BG403555 602419260
44	669.5	27.0	788	10 BG163397	BG163397 602338360
45	663	26.7	481	10 BF729761	BF729761 mab75g12.

#### ALIGNMENTS

RESULT 1  
LOCUS AK016546 3143 bp mRNA linear HTC 05-DEC-2002  
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog  
[Homo sapiens], full insert sequence.  
ACCESSION AK016546  
VERSION AK016546.1 GI:12855334  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1	Carninci, P. and Hayashizaki, Y.
AUTHORS		
TITLE		High-efficiency full-length cDNA cloning
JOURNAL		Methods. Enzymol. 303, 19-44 (1999)
MDLINE		99279253
PUBMED		10349636
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
AUTHORS		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
TITLE		Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL		
MDLINE		20499374
PUBMED		11042159
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nageoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, K., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kir, A. and Hayashizaki, Y.
AUTHORS		RIKEN integrated sequence analysis (RISA) system-384 format
TITLE		sequencing pipeline with 384 multichipillary sequencer
JOURNAL		Genome Res. 10 (11), 1757-1771 (2000)
MDLINE		20530913
PUBMED		11076861
REFERENCE	4	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arai, K., T. Hara, A., Fukunishi, Y., Kono, H., Aachi, T., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadote, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Knehl, P., Lewis, S., Matsuo, Y., Nakado, I., Penzo, G., Quackenbush, J., Schriml, L.M., Strubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Baren, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Galardi, M., Gustincich, S., Hill, D., Hochmann, L., Hume, D.A., Kamuya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, R., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyono, O., K. Wang, K.H., Weitz, C., Whiteaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Haegawa, Y., Kawai, H., Kohitsuki, S. and Hayashizaki, Y.
AUTHORS		Functional annotation of a full-length mouse cDNA collection
TITLE		Nature 409 (6821), 685-690 (2001)
JOURNAL		
MDLINE		21085660
PUBMED		11217851
REFERENCE	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
AUTHORS		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
TITLE		Nature 420, 563-573 (2002)
JOURNAL		
MDLINE		6 (bases 1 to 3143)
PUBMED		
REFERENCE		Aachis, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arai, K., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
AUTHORS		Direct Submission
TITLE		Submitted (10-JUL-2000)
JOURNAL		Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9216) Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTTAAATTAATATCCCTCCCCCCC 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAAATTAATATCCCTCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size compised longer than 7 kb was selected before cloning. Vector: a modified plusscript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

FEATURES

Source

Location/Qualifiers

1..3143

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:4932434F09"

/db\_xref="MG1:1907554"

/dd\_xref="taxon:10090"

/clone="4932434F09"

/sex="male"

/tissue\_type="testis"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

327..3005

/note="unnamed protein product; DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens] (SPR1|Q9HNM5, evidence: FASTY, 95.7%id, 100%length, match=2649)

putative"

/codon\_start=1

/protein\_id="BA830295.2"

/db\_xref="GI:26382128"

/dd\_xref="MG1:1921638"

/translation="MKIPSGRCMGMAAMETEOUGVEIFETACEGNGESQDRPKLEP FYEYRWSQLKLADLTRKHGYMMAKAPHDMPVKTPTDPSHSDRVYVYLAMSGEN RENTLFESELPKTNRAAVYLMISMKPLDLFOATLDGMYREELREKRIGTGV1 AADVYHSGSTFLFOAGSGIVYIKDGGPHGFTOOLRPNVVENSCPNIRMDPKLCPAD PMDIATHSNDITVSNLTVREERITVYNNELNAMESDPASACVATFVLOEEDRYSG YMCPDAERTPSGGKILRIIYENDSESEVETHTVSMLETRADSRFRYKGTANKK VTRKSEIVVDAAGI1DIVIDKLVQPFELIFEGVETIAGAGTPEGRKAWSLTLDSS QHTLQVLSPELIFVEDDAMDRORLIESVPDSVPLIYEBTDDIMINHIHFVF TSEEMVLRHGSINIVWDERKLVVEGTDSPLDEHNLVTSYANPEVVRVLDRGYS HSCCLSRHGDPEFISKYSKNKPHCVSLYKLSPEDDPVHKTKFPATILLASGLPLPY TPPEITSFESTTGFTLYGMKXPHDQPKKTYTVLYGGPOVLYNNRFKGYKTR LNTLASGVIVVYIDNRGSHRGLKTEGAKYMGQIEIDDVYEGLYVLA5YDPDL DRYGIVAMAOAEKRPSEPNRLLLHGFILBNVFAHTSILSLVYBAGKPYDQIVYQ ERHSIVPESGEHYELHLHYLOENGSRIALAKVI"

BASE COUNT

894 a 688 c 773 g 788 t

ORIGIN

Alignment Scores:

Primed. No.:

Score: 1.37e-245 Length: 3143

Percent Similarity: 2326.00 Matches: 443

Best Local Similarity: 80.18% Conservative: 10

Query Match: 78.41% Mismatches: 12

93.71% Indels: 100

11 Gaps: 1

US-10-070-464-5 (1-465) x AK016546 (1-3143)

```

Db      1308  AAGGACAGCAAAACCAAGGCACTTCAAGATGTCGAGATTGTGTGATGCTGA 1367
Qy      21  G|A|g|l|e|l|e|a|s|p|V|a|l|l|e|a|s|p|y|g|l|u|e|l|e|g|l|n|P|r|o|p|h|e|g|l|u|e|l|e|u|p|h|e|g|l|u| 40
Db      1368  GAGGAGATTATGATGTCATAGATAAAGAACTGGTTCACCTTCCAGATTCTGTTGAG 1427
Qy      41  G|V|a|l|g|l|u|T|y|l|e|l|a|r|g|a|l|a|g|l|T|P|T|P|P|r|o|g|l|u|g|l|y|s|e|T|y|l|a|T|P|S|e|l|l|e| 60
Db      1428  GAGATTGAATATATATGTCGAGCGGAGTGGATCTCCAGAGGAGAAACATCCCTGGTCCATC 1487
Qy      61  L|e|u|e|a|s|p|a|s|e|r|g|l|n|t|H|a|r|g|l|e|u|g|l|l|e|a|l|l|e|s|e|r|P|r|o|g|l|u|e|u|p|h|e|l|e| 80
Db      1488  C|T|A|C|T|A|G|A|C|G|T|T|C|C|A|G|A|C|T|C|G|A|G|A|T|A|G|T|T|C|T|G|A|T|T|C|C|C|T|G|A|T|T|A|T|T|C|C| 1547
Qy      81  P|r|o|v|a|l|g|l|u|a|s|p|a|V|a|l|e|t|e|l|u|a|g|l|n|a|g|l|e|u|l|e|g|l|u|s|e|r|V|a|l|P|r|o|a|s|p|e|r|V|a|l| 100
Db      1548  C|C|A|G|A|T|A|G|A|G|A|T|G|C|C|A|T|G|A|C|A|G|A|G|A|C|T|T|A|G|A|G|C|A|G|T|T|C|C|T|G|A|C|T|C|T|G|G| 1607
Qy      101  T|H|P|P|r|o|l|e|u|l|e|l|e|T|y|g|l|u|g|l|u|T|H|P|T|P|P|l|e|T|P|l|e|a|s|n|l|e|H|s|a|s|p|l|e|P|H|e| 120
Db      1608  A|C|A|C|C|A|G|A|T|C|A|T|C|A|T|A|G|A|A|A|C|A|C|A|G|A|C|A|T|G|A|T|T|A|T|C|C|A|G|A|T|T|T|T| 1667
Qy      121  H|I|s|V|a|l|P|H|e|P|r|o|g|l|n|S|e|r|H|s|g|l|u|g|l|u|l|e|g|l|u|P|H|e|l|e|P|H|e|a|l|S|e|r|g|l|u|C|y|s|l|y|s| 140
Db      1668  C|A|T|G|T|T|T|T|C|C|C|A|A|C|T|C|A|G|A|G|A|T|G|A|G|T|T|A|T|T|T|T|G|C|T|G|A|T|G|C|A|A| 1727
Qy      141  T|H|g|l|y|P|H|e|A|r|g|H|s|l|e|u|T|y|l|y|s|l|e|T|H|S|e|r|l|e|l|e|u|l|y|g|l|u|S|e|r|l|y|T|y|l|y|S|a|r|g| 160
Db      1728  A|C|A|G|G|T|T|T|C|G|A|C|A|C|T|G|A|T|A|A|A|C|A|C|A|T|C|A|T|T|T|A|A|G|A|G|C|A|A|A|T|A|A|A|C|G|G| 1787
Qy      161  S|e|r|S|e|r|g|l|y|l|e|u|P|r|o|l|a|P|r|o|S|e|r|A|s|P|H|e|l|y|S|e|r|P|r|o|l|e|y|g|l|u|l|e|a|l| 180
Db      1788  T|C|C|A|G|T|G|G|A|C|T|A|C|T|C|C|C|C|A|G|A|G|T|T|T|C|A|G|T|G|C|T|A|T|C|A|A|A|A|G|A|A|T|A|A|C|A| 1847
Qy      181  I|l|e|T|H|S|e|r|g|l|y|l|e|u|T|P|g|l|u|a|l|e|u|g|l|y|a|r|g|H|s|g|l|y|S|e|r|a|s|n|l|e|g|l|n|V|a|l|S|p|g|l|u| 200
Db      1848  A|T|T|A|C|C|A|G|G|G|T|G|A|A|T|G|G|A|G|A|C|T|T|G|C|C|G|C|G|A|G|A|T|C|T|A|T|A|T|C|T|G|G|G|T|G|A|T|G|A| 1907
Qy      201  V|a|l|a|g|a|g|l|e|u|V|a|l|T|y|r|P|H|e|g|l|u|g|l|T|H|T|y|l|S|a|s|P|S|e|r|P|r|o|l|e|u|H|I|S|I|S|l|e|u|T|y|r| 220
Db      1908  G|C|C|A|A|A|A|G|C|T|G|T|G|T|A|C|T|T|T|A|A|G|G|C|A|A|A|G|A|C|T|C|C|T|T|G|G|A|G|A|C|A|C|C|T|G|T|A|C| 1967
Qy      221  V|a|l|S|e|r|T|y|r|V|a|l|S|a|n|P|r|o|g|l|y|g|l|u|a|l|T|H|a|r|g|l|e|u|T|H|A|s|P|a|r|g|l|y|T|y|S|e|r|H|s| 240
Db      1968  G|T|G|A|C|C|A|G|T|T|A|T|G|C|A|A|A|C|C|C|T|G|A|G|A|G|T|G|G|A|G|G|C|T|A|C|T|G|A|C|C|G|T|G|C|T|A|C|C|A|C| 2027
Qy      241  S|e|r|C|y|s|l|e|S|e|r|g|l|n|H|s|C|y|s|a|s|P|H|e|P|H|e|l|e|S|e|r|l|y|T|y|S|e|r|a|s|n|g|l|u|y|S|a|n| 260
Db      2028  T|C|C|T|G|C|T|G|C|C|A|G|C|G|C|G|A|T|T|G|A|C|T|T|C|T|A|T|A|G|A|T|A|G|A|C|A|C|A|G|A|A|G|A|T| 2087
Qy      261  P|r|o|H|s|C|y|s|V|a|l|S|e|r|l|e|u|T|y|r|l|y|S|e|u|S|e|r|S|e|r|P|r|o|g|l|u|a|s|P|P|r|o|T|H|C|y|s|l|y|S|H|r| 280
Db      2088  C|C|A|C|A|C|T|G|T|G|T|C|C|C|T|A|C|A|A|A|C|C|T|C|A|G|T|C|C|T|G|A|G|A|G|A|G|A|C|C|A|G|T|C|A|T|A|A|A|C|A| 2147
Qy      281  L|y|S|g|l|u|P|H|e|T|P|A|l|A|T|H|T|l|e|u|S|P|S|e|r|a|l|a|g|l|P|r|o|l|e|u|P|r|o|S|P|T|y|r|H|P|r|o| 300
Db      2148  A|A|G|A|A|T|T|T|T|G|G|C|C|A|C|A|T|T|T|G|A|T|T|A|G|A|G|G|T|C|C|T|C|C|G|A|C|T|A|C|C|C|C|T|C|A| 2207
Qy      301  G|l|u|l|e|P|S|e|r|P|H|e|g|l|u|S|e|r|T|H|T|g|l|y|P|H|e|r|l|e|u|T|y|r|g|l|M|e|t|l|e|u|T|y|r|l|y|S|P|r|o| 320
Db      2208  G|A|A|T|T|T|T|T|T|T|T|T|G|A|A|G|T|A|C|T|G|A|T|T|T|A|C|T|G|T|A|G|A|T|G|T|T|G|T|A|A|G|C|T| 2267
Qy      321  H|I|s|a|s|P|l|e|u|g|l|n|P|r|o|g|l|y|l|y|S|y|T|y|r|P|r|o|T|H|V|a|l|l|e|u|P|H|e|l|e|T|y|r|g|l|y|g|l|u|P|r|o|g|l|n| 340
Db      2268  C|A|T|G|C|T|A|C|A|C|C|T|G|A|A|A|A|A|T|A|C|C|C|A|C|T|G|T|G|T|A|T|T|A|T|G|A|T|A|T|G|T|G|T|C|C|C|A|G| 2327
Qy      340  ----- 340
Db      2328  G|T|G|C|A|G|C|T|G|T|G|A|C|A|T|C|G|T|T|A|A|G|A|G|T|C|A|A|G|T|A|T|T|C|C|G|C|T|G|A|A|C|C|C|T|G|G|C| 2387
Qy      340  ----- 340

```

```

Db      2388  TCCCTGGTATTATGTTGTTGTGTGATAGACAAAGGGGATCCGTGCACCGAGACTTAA 2447
Qy      340  ----- 340
Db      2448  TTTGAAGGCGCTTTAAATATATAAAATGGTCAAAATAGAAATGATGATCAAGTGAAGA 2507
Qy      340  ----- 340
Db      2508  CTCGAGTACCTTAGCATCTCAGTATAGACTTCAATTGACTTGGATCGAGTGGGATCCAGGC 2567
Qy      340  ----- 340
Db      2568  TGGTCTATAGTGGTACTCTCCCTGATGGCATTAAATGACAGAGTGGATATCTTCCGG 2627
Qy      341  V|a|l|a|l|e|a|l|g|l|A|P|r|o|V|a|l|H|r|l|e|u|T|P|l|e|P|H|e|T|y|r|A|s|P|T|H|g|l|y|T|H|g|l|u| 360
Db      2628  G|T|G|G|C|T|A|T|T|G|C|T|G|G|G|G|C|C|C|C|A|C|C|C|G|T|G|A|C|T|T|C|T|A|T|A|T|A|C|A|G|A|T|A|C|A|G|G|A| 2687
Qy      361  A|r|g|T|y|r|M|e|r|G|l|H|s|P|r|o|S|p|l|n|a|n|g|l|u|g|l|n|g|l|T|y|r|T|y|r|l|e|u|g|l|S|e|r|V|a|l|A|M|e|t| 380
Db      2688  C|C|C|T|A|T|A|T|G|G|T|C|A|C|C|C|T|G|A|C|C|A|G|A|T|G|A|A|C|G|G|C|T|A|C|T|A|G|A|T|C|T|G|G|C|A|T|G| 2747
Qy      381  G|l|n|a|l|a|g|l|u|y|P|H|e|P|r|o|S|e|r|g|l|u|P|r|o|S|a|n|a|r|g|l|e|u|l|e|u|l|e|u|H|s|g|l|y|P|H|e|u|A|S|P| 400
Db      2748  C|A|G|G|G|A|G|A|A|T|T|C|C|C|C|C|A|G|A|C|A|A|C|G|G|T|A|C|T|C|T|T|G|C|A|T|G|G|T|T|C|T|T|G|A|T| 2807
Qy      401  G|l|u|a|n|V|a|l|H|s|P|H|e|A|l|H|s|T|H|S|e|r|l|e|l|e|u|l|e|u|S|e|r|P|H|e|u|V|a|l|r|g|a|l|y|l|y|S| 420
Db      2808  G|A|G|A|T|G|T|C|A|C|T|T|G|A|C|A|C|C|A|G|A|T|A|T|G|C|A|G|A|G|T|T|T|T|T|G|A|G|G|C|T|G|G|A|A|G| 2867
Qy      421  P|r|o|T|y|r|A|S|P|l|e|u|g|l|l|e|T|y|r|P|r|o|g|l|n|u|a|r|g|H|s|S|e|r|l|e|a|r|V|a|l|P|r|o|g|l|u|S|e|r|g|l|y| 440
Db      2868  C|C|A|T|A|G|A|C|T|T|A|C|A|T|C|A|T|C|T|C|A|G|A|G|A|G|C|A|C|A|G|A|T|A|G|G|T|T|T|T|G|A|G|G|C|T|G|G|A| 2927
Qy      441  G|l|u|H|s|T|y|r|G|l|u|e|u|H|s|l|e|u|e|u|H|s|T|y|r|l|e|u|g|l|n|g|l|u|a|s|n|l|e|g|l|y|S|e|r|a|r|g|l|e|a|l| 460
Db      2928  G|A|A|C|A|C|T|A|T|G|A|C|T|C|A|C|C|T|G|C|A|C|T|T|C|A|G|A|G|A|C|T|T|G|A|G|A|C|T|T|G|A|T|C|G|T| 2987
Qy      461  A|l|a|l|e|u|y|V|a|l|l|e| 465
Db      2988  G|C|T|G|A|A|A|G|T|G|A|T| 3002

```

RESULT 2  
 AK029788  
 LOCUS  
 DEFINITION  
 Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930560C15 product:DIPEPTIDYL PEPTIDASE 8 homolog (Homo sapiens), full insert sequence.  
 AK029788  
 AK029788.1 GI:26081520  
 VERSION  
 KEYWORDS  
 HTG; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL  
 MEDLINE  
 PUBMED  
 11042159  
 20499374  
 11042159

REFERENCES  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashtwari, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watehiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)

```

QY      262  HiscysValSerLeuTyrIleuSerSerProGluuAspProThrCysIleThrIys 281
          |||||
Db      3923  CACTGTGTGTCCTCTCAAACTCTCAAGTCTAGAGATGACCACTTATATAAACAAG 3982
QY      282  GluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGlu 301
          3983  GAATTTTGGGCGACCATTTTGGATTGACAGGTCTCTTCTGACTACACCCCTCCAGAA 4042
QY      302  IlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHis 321
          4043  ATTTTTCCTTTTGAAGTACTACTGAGATTACACTGATGGAATGTGTATTAACCTCAT 4102
QY      322  AspLeuGlnProGlyIleValSerTyrProThrValLeuPheIleTyrGlyIleProGln 340
          4103  GACCTACAACTGGAAAGAAATACCCACCTGTGTTATTCATATATGGTGGTCCACAGTG 4162
QY      340  -----
          4163  CAGCTGTGAACAATCGTTTAAAGAGTCAAGTATTTCCGCCCTGAACACCTGGCTCC 4222
QY      340  -----
          4223  CTGGGTTATGTGTGTGTGATAGACACAGGGGATCTGTCAACGAGACTTAATTT 4282
QY      340  -----
          4283  GAAGCGCCCTTAAATATTAATATGCGTCAAAATAGAAATCGATGATCAAGTGAAGAGCTC 4342
QY      340  -----
          4343  CAGTACTAGCATCTAGTATGACTTCACTTGGATGGAGTGGGATCCACGGCTGG 4402
QY      341  -----Val 341
          4403  TCCTATGTGGTCACTCTCCCTGATGACATTAATGACAGAGTGGATATCTTCCGGGTG 4462
QY      342  AlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArg 361
          4463  GCTATTGCTGGGGCCCGACGACACCTCTGATCTCTTATATACAGATACACGAGACGC 4522
QY      362  TyrMetGlyIleProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGln 381
          4523  TATATGGTCACTCCCTGACCAAGATGAAACAGGGCTACTACTAGGATCTGTGGCATCA 4582
QY      382  AlaGluIlePheProSerGluProAsnArgLeuLeuLeuIleGlyPheLeuAspGln 401
          4583  GCGGAGAGTTCCTCCCTGAGAACCAACCGGTACTCTCTTGCATGGGTCTTGGATGAG 4642
QY      402  AsnValIlePheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIlePro 421
          4643  AATGTTCACTTTCACACACACCATATATTCCTGAGTTTTTTAGTGAAGGCTGGAAAGCCA 4702
QY      422  TyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProIleSerGlyGlu 441
          4703  TATGACTTACAGATCTATCTCTCAGAGAGGACACAGCATCAGAGTTCTGTGAGGAA 4762
QY      442  HisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 461
          4763  CACTATGAACTGACACCTGCTCCACTACTTACAGAGAACCTTGATGCGTATGCGTCT 4822
QY      462  LeuIleValIle 465
          4823  CTCGAAAGTGTATA 4834
Db      4823  CTCGAAAGTGTATA 4834

```

```

RESULT 3
LOCUS    AK050021
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length
           enriched library, clone: C730003D12 product: D1PERP1D1L, PERP1DASE 9
           homolog [Homo sapiens], full insert sequence.
ACCESSION AK050021
VERSION    AK050021.1 GI:26340743
KEYWORDS   HTC; CAP trapper.

```

```

SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS  Carninci, P. and Hayashizaki, Y.
TITLE    High-efficiency full-length cDNA cloning
JOURNAL  Meth. Enzymol. 303, 19-44 (1999)
MEDLINE  99279253
PUBMED   10349636
REFERENCE 2
AUTHORS  Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
          Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE    Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL  Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE  20499374
PUBMED   11042159
REFERENCE 3
AUTHORS  Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
          Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
          Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
          Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
          Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
          Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
          Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE    RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer
JOURNAL  Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE  20530913
PUBMED   11076861
REFERENCE 4
AUTHORS  Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
          Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
          Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakawa, I.,
          Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R.,
          Kadota, K., Matsumoto, H., Ashburner, M., Batalov, S., Casavant, T.,
          Frieschmann, W., Gaasterland, T., Glass, C., King, B., Kochiwa, H.,
          Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
          Quackenbush, J., Schriml, L.M., Staahl, F., Suzuki, R., Tomita, M.,
          Wagner, L., Washio, T., Sakai, K., Okado, T., Furuno, M., Aono, H.,
          Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
          Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
          Fletcher, C., Fujita, M., Gariboldi, M., Gietzung, S., Hill, D.,
          Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
          Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,
          Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
          Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
          Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
          Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S.
          and Hayashizaki, Y.
TITLE    Functional annotation of a full-length mouse cDNA collection
JOURNAL  Nature 409 (6821), 685-690 (2001)
MEDLINE  21085660
PUBMED   11217851
REFERENCE 5
AUTHORS  The PANTOM Consortium and the RIKEN Genome Exploration Research
          Group Phase I & II Team.
TITLE    Analysis of the mouse transcriptome based on functional annotation
          of 60,770 full-length cDNAs
JOURNAL  Nature 420, 563-573 (2002)
PUBMED   12108560

```





QY	340	-----	340
Db	2234	CACCTTCAGAGGGGCGCCCTGAAAAATGCAATGGCCAGGTGAGACATTGACGACCGAGGTGAA	2293
QY	340	-----	340
Db	2294	GGCTTGACGTAGTGGCTGAGAAAGTATGGCTTCATTGACTTGAGCCGAGTGGCCATCCAT	2355
QY	340	-----	340
Db	2354	GGCTGGTCTTACGCGGGCTTCTCTCACTCATGCGGCTCATTCACAAAGCTGTTTC	2413
QY	341	---ValAlaIleAlaGlyValAProValThrLeuTrpIlePheTyrAspThrGlyTyrThr	359
Db	2414	AAGGTAGCCATTGGCGGGCGCTCTCTGCTACTGTGTGAATGGCCTTATGACACAGGGTACACG	2473
QY	350	GIuArGTyMeGlyNHisProAspGlnAsnGlnGlnGlyTyrTyrTyrLeuGlySerValAla	379
Db	2474	GAAAGCATGATGATGATGTCGCCGGAATAAATTAACAGCAAGGCTATGAGGACGAGGTGTGTACCC	2533
QY	380	MetGlnAlaGlyuPhePheProSerGluuPheAsnArgLeuLeuLeuLeuHisGlyPheLeu	399
Db	2534	CTGCATGTGGAGAACTGCCATGCCAATGAGCTTAACCCCTGCTTATCTCTCACGGCTTCTCGT	2593
QY	400	AspGluAsnValHisPheAlaHisThrSerIleuLeuSerPheLeuValArgAlaGly	419
Db	2554	GACGAGAAAGCTTCACTTCTTCCACACAAATTTCTCGTGTGCCAGCTATTCGAGACAGGA	2653
QY	420	LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer	439
Db	2654	AAGCCATACCAAGCTTCAGATCTACCCAAACGAGACATACATCCGCTGCGGAGATCC	2713
QY	440	GlyGluHisTyrGlnLeuHisLeuLeuHisTyrIleuGlnGlnGluAsnLeu	455
Db	2714	G3AGAGCATTTACGAGGTGACGCTGCTGCACTTCTTGACAGAAACCTTG	2761
RESULT 4			
AK050023			
LOCUS			
DEFINITION	AK050023	3376 bp	mRNA
			linear, HTC 05-DEC-2002
			Mus musculus adult male liver tumor cDNA, RIKEN full-length
			enriched library, clone: C730030H19 product: D18EPDIT1 PEPDITASE 9
			homolog [Homo sapiens], full insert sequence.
ACCESSION	AK050023		
VERSION	AK050023.1	GI:26093756	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
TITLE	Ichih, M., Komuro, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
	Normalization and subtraction of cap-trapper-selected cDNAs to		
	prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE			
AUTHORS	Shibata, K., Ichih, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,		
	Komuro, H., Akiyama, J., Nishi, K., Kitamura, T., Tashiro, H., Ichih, M.,		
	Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimi, T., Harada, A.,		
	Yamanoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,		
	Fujiwara, R., Inoue, K., Togawa, Y., Irawa, M., Ohara, E., Matsubuchi, M.,		
	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, U.,		
	Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.		
	RIKEN Integrated sequence analysis (RISA) system--384-format		
TITLE			

JOURNAL  
MEDLINE  
PUBMED  
REFERENCES  
AUTHORS

4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kaubakawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H., Kuehl, P., Lewis, S., Matsumoto, Y., Nakai, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabile, F., Suzuki, R., Tomita, M., Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bul, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hoffmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Togo, Oka, K., Wang, K.H., Wetz, C., Whitaker, C., Wilming, L., Wymshar-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohatsu, S., and Hayashizaki, Y.

JOURNAL  
MEDLINE  
PUBMED  
REFERENCES  
AUTHORS

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3376)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Furuda, S., Furuno, M., Hangaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imochi, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Okachi, M., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsumura, T., Miyazaki, A., Murata, M., Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohmoto, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sato, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashi, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

JOURNAL  
MEDLINE  
PUBMED  
REFERENCES  
AUTHORS

1  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers

FEATURES  
SOURCE

1. 3376  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:C730003H19"

```

/db_xref="taxon:10090"
/clone="C730003H19"
/sex="male"
/cell_type="tumor"
/tissue_type="liver"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
235..2824
/notice="DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens)
(SPTKRAL47179, evidence: FASTA, 92%ID, 99.6%length,
match=2580)
misc_feature

```

BASE COUNT 698 a 1046 c 932 g 700 t

## Alignment Scores:

```

Pred. No.: 7.4e-153 Length: 3376
Score: 1488.50 Matches: 278
Percent Similarity: 62.59% Conservative: 70
Best Local Similarity: 50.00% Mismatches: 107
Query Match: 59.97% Indels: 101
DB: 11 Gaps: 2

```

US-10-070-464-5 (1-465) x AK050023 (1-3376)

```

QY 1 ThrGlyThrAlaAspProlyValThrPheLysMetSerGluLeuMetIleAspAlaGlu 20
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1154 ACAGGCGACGAAAGAACCCCAAGATTGCCGTGAGCTGAGCTCCAGACGACATCAG 1213
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1214 GGCMAAATCGTGTCAAGCTGAGAGAGAACTGTGTACAGCATTCACCTCTTTTCCC 1273
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1274 AAAGTGAAGTACATCGCCCGGCTGCTGAGACACGGACCGCAATATGCTGGGCGCATG 1333
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1334 TTCCTGGACCGTCCCGACGAAACGGCTTCAGCTTGTCTCTGCGCCCTGCTCTTCATC 1393
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 81 ProValGluAspArgValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1394 CCGGCGCTTGAAGTGAAGGCCAGGCGAGCAGCTGCCAGAGCGTCCCAAGAAATGTG 1453
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1454 CAGCCCTTGTGATCTATGAAGAAGTCAACATGTCTGATCAACGTCCACGACATCTTC 1513
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 121 HisValPheProGlnSerHis--GluGluGluIleGluPheIlePheIleAspGluCys 139
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1514 CACCGCTTCTCAGGCTGAGGCGCAGACAGACTTTGTTTCTTCGTGCAACGAATGC 1573
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 140 LysThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1574 AAAGTGTGCTTGTGCCACCTGTACAGGCTCAAGTGAACCTTAAACCAAGCACTAGAC 1633
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 160 ArgSerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIle 179
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1634 TGAGCGGAACCCCTCAGCCCTACAGAAAGATGATTAAAGTCCCAACAAAGAGAGGTC 1693
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 180 AlaIleThrSerGlyLysTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1694 GCGCTGACGATGCGGCGGTGTCTTCGAGGATGATGATCAAGATCTGGGCTCAAC 1753
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 200 GlyValArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeu 219
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1754 GAGACGACGAAGCTGTGATCTTCAAGGTAACAAAGACACCGCTGGAACATCACCTC 1813
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 220 TyrValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSer 239
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1814 TATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1873
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

QY 240 HisSerCysValIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1874 CACAGGTGCTCCATGAGCCAGAGCTTCGACATGTTGCTGAGTCACTACACAGCTGAGC 1933
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 260 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 279
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1934 ACGCCACCTGTGTATCATGTGTCAAGCTGAGGCGGCCCATATATGACCACTGCACAG 1993
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 280 ThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 299
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1994 CAACACGACTTCTGGGCGACATGATGAGGCGACCAATGCCCCCAAGCTATGTGCC 2053
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 300 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 319
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2054 CTTGAGATCTTCCATCTCCACCGCCAGATGATGAGGCGACGTCGAGCTTACGGCATGATCAAG 2113
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 320 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 339
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2114 CCACACACCTGTGAACTGGAGGAGGACCCCACTGTGCTTGTGTATGGGCGCCA 2173
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 340 Gln----- 340
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2174 CAGGTGATGTGGTGAACAACCTCTTAAGGATCAAAATCCTCGCGCTAAATACACTG 2233
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 340 ----- 340
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2234 GCATCTTGGGCTATGCTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2293
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 340 ----- 340
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2294 CACTTGAAGGAGGCGCTGAAATAATCAATGGCCAGGTGAGATTGAGACCAAGTGA 2353
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 340 ----- 340
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2354 GCGTTCAGTACGTGGCTGAGAAATGATGCTTATTGACTTGAAGCCAGTCCCATCCAT 2413
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 340 ----- 340
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2414 GCGTGTCTTACGCGGCTTCTCTCATCATGATGATGATGATGATGATGATGATGATGATGAT 2473
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 341 ---ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThr 359
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2474 AAGGTAGCCATTCGAGGCGCTCTGCTGATGATGATGATGATGATGATGATGATGATGAT 2533
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 360 GluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlnSerValAla 379
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2534 GAACGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2593
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 380 MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2594 CTGCATGTGAGAAAGCTGCCCAATGAGCTTAAACCGCTGCTTATCTCCACAGGCTTCCTG 2653
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGly 419
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2654 GACGAAACCTTACTTCTTCCACACAAATTTCTGTGTGTCAGCTGATGCAAGACAGA 2713
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2714 AACCATACAGAGCTTACATGATCTACCAACAAAGAGACATAGATCCGCTGCCGAGTCC 2773
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeu 455
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2774 GGAGACCATTAACGAGGTGACGCTGTGCACTTCTCAGAAACACTG 2821
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

## RESULT 5

AK078301 3457 bp mRNA linear HTC 05-DEC-2002  
 LOCUS AK078301  
 DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
 enriched library, clone:6430584G1 product:DIPEPTIDYL PEPTIDASE 9  
 homolog (Homo sapiens), full insert sequence.  
 ACCESSION AK078301  
 VERSION AK078301.1 GI:26347124  
 KEYWORDS HTC; CAP trapper.

SOURCE ORGANISM	Mus musculus (house mouse)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	Carninci, P., and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	103496636
2	
REFERENCE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	20499374
REFERENCE	11042159
3	
REFERENCE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, Y., Nishi, K., Kutsuna, T., Teshito, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	11076861
4	
REFERENCE	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kaenkawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakai, I., Pesole, G., Quackenbush, J., Schriml, L. M., Skidell, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Hoffell, D., Bojunga, N., Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohzuki, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
PUBMED	21085660
REFERENCE	11217851
5	
REFERENCE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL	Nature 420, 563-573 (2002)
PUBMED	6 (bases 1 to 3457)
REFERENCE	Adachi, J., Aizawa, K., Akiyama, T., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, P., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanabe, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

FEATURES	source
COMMENT	<p>Muramatsu,M. and Hayashizaki,Y.  Direct Submission  Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Physical and Chemical Research, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Saitoro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan [E-mail:genome-resgscgsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9216, Fax:81-45-503-9216]  cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  Please visit our web site for further details.  URL:http://genome.gsc.riken.go.jp/  URL:http://fantom.gsc.riken.go.jp/  Location/Qualifiers</p>
FEATURES	<p>1..3457  /organism="Mus musculus"  /mol_type="mRNA"  /strain="C57BL/6J"  /db_xref="PANTOM:DB:6430584G11"  /db_xref="taxon:10090"  /clone="6430584G11"  /sex="male"  /tissue_type="olfactory brain"  /clone_id="RIKEN full-length enriched mouse cDNA library  /dev_stage="adult"  182..2770  /home="unnamed protein product; Dipeptidyl peptidase 9 homolog [Homo sapiens] (SPRT:AA147179, evidence:FASTA, 92%ID, 99.6%length, match=2580)  putative"  /codon_start=1  /protein_id="BAC37211.1"  /db_xref="GI:26347125"  /translation="MCSGVSPVEQVAAAGMDMDTAARFCVQKSHMDGLRSIIHGSRKSGLTYSAKHDFQFOVKPDSPGSHRLTYLGMPSRSRELSLSEIPIKVRKEALLSWKMDLHDFQATPHHGYVREBELREKRRLGFGITSDHSEGLFLFOASNLFLCRDGGKGFMYSPMKPLEIKTQCGSPDRPKICPADPAFSTINNSDLVANIETGERRLTFCQHGSGAGVDNFKSAGVATFVIOEDDFRTGCMWCPYASWBSGLTLTRLIEVESSEVIVHPSPALBERTDSYPRGSKNPKALKAELOTHQKIVSSCKELVQFSSLPKVEYIARACWTRGKAMAMFLRPOORLOLVLLPALFIPAVESAAORAAAVAVKNOQFVITYEIVTNVIMVHDIFHPFOAEQDQFCFLRNECKTGCHLVRVAVELTKDYDWTDEPLSPEDDEFKCPKIEKVALTSGMEVLSRGSKIWNBE TKLYVFOSTKDPRLAEHLIYVSYESAGIYVLTITGFSHSCMSGSPMFVHSYSVTPCVHYTKLSGPDDDLKQPRFMAWMEANCPDVVPEIPIHRTADVLQGM YKPHLQPRGHFVLLFYGGPQVQVUNNSKGIYDLRLANTLSGYAVAVIDGRSSQRGHAFEGALKNQMGQVEIEDQVGLQVAAEYGFIDLSRVAIHGMSSYGGFSLMGLHKQPFKAVIAGAPVTWMAVDVGTGYEMVDPENKKGQYEGASVALAHKXLPNEPNI LLIHGFLEVDVHFHTNIFLVSQLIRACKPYQLQIYPRNBSIRRGSEHIEVTLILLFLQSHL"</p>
BASE COUNT	766 a 1033 c 926 g 732 t
ORIGIN	
Alignment Scores:	
Pred. No.:	9.82e-152 Length: 3457
Score:	1478.50 Matches: 277
Percent Similarity:	62.41% Conservative: 70
Best Local Similarity:	49.82% Mismatches: 108
Query Match:	59.57% Indels: 101
Gaps:	11 2
US-10-070-464-5 (1-465) x AK078301 (1-3457)	
0y	1 ThrGlyThralaaanProLyValThrPhelYsmetSerGiulilemetIleaspaIagu 20
Db	1100 ACAGCGACGAAGAACCCCAAGATTGGCCCTGAGAGCTGAGCTCCACGACGACATCAG 1155
0y	21 GlyArlleleleasValIleasPlygIuleuleginPropheglurIleuaphegu 40
Db	1160 GGCMAAATCGGTCAAGCTCGACGAGGAACATGGTACAGCCATTCAGCTCCCTTTTCCC 1219

QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrPheProGluGlyLysTyrValAspSerIle 60  
 |||||.....  
 Db 1220 AAGGTGAGTACATCCCGCGGCTGGCTGGACACGGGACGCAAAATATGCTGGCGCATG 1279  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 |||||.....  
 Db 1280 TTCCTTACCGTCCCGCAACCGCTTCAGCTTCTCTCCCTACCCCTGCTCTCTTCATC 1339  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 |||||.....  
 Db 1340 CCGGCGGTTGAGGTGAGGCGCCAGCGGACGCTGCCAGACCGTCCCGCAAGATGTG 1399  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 |||||.....  
 Db 1400 CAGCCTTTGTCTCTATGAGAAGTACCAATGTCTGATCAACGTCCAGCATCTTC 1459  
 QY 121 HisValPheProGlnSerHis---GluGluGluIleGluPheIlePheAlaSerGluCys 139  
 |||||.....  
 Db 1460 CACCCGTTCTCAGGCTGAGGCGCAGCAGACTTTGTTCTCTGCTGCGCAAGATGC 1519  
 QY 140 LysThrGluPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159  
 |||||.....  
 Db 1520 AAGACTGGCTTCTGCCACCTGTACAGAGGTACAGTGAACTTAAACCAAGACTATGAC 1579  
 QY 160 ArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIle 179  
 |||||.....  
 Db 1580 TGGACCGAACCCTCAGCCCTACAGAAAGATTTAAGTCCCGCCATCAAGAGAGAGTCC 1639  
 QY 180 AlaIleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199  
 |||||.....  
 Db 1640 GCCCTGACAGGTGGCAAGTGGAGGAGTCTGTCTGCAAGGATGGCTCAAGATCTGGGTCAAC 1699  
 QY 200 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIle 219  
 |||||.....  
 Db 1700 GAGCACAAGAGTGTGTGTTCTTCAAGGTACAAAGACACACGCTGGAACTACCTC 1759  
 QY 220 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 239  
 |||||.....  
 Db 1760 TATGTGTGCTAGTACAGTACAGCAGAGGAGATGCTGGCTGCTCACACGCTCGGCTCTCC 1819  
 QY 240 HisSerCysAlaSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259  
 |||||.....  
 Db 1820 CACAGTGTCTCCATGAGCCAGACCTTGCACATGTTCTGTGATCCTACACACGACTGTGAGC 1879  
 QY 260 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 279  
 |||||.....  
 Db 1880 ACGCCACCTGTGTACGTGTACAGTACAGTACAGCGCCCGCATATGACCCACATGCACAG 1939  
 QY 280 ThrLysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 299  
 |||||.....  
 Db 1940 CAACACAGCTTCTGGGCGACAGATGATGAGAGCAGCCAAATGCCCCAGACTATGAGGCC 1999  
 QY 300 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 319  
 |||||.....  
 Db 2000 CCGAATCTTCCATCTCCACACCGCTGACAGCTGACGTGACGTACCGCATGATCTCTCAAG 2059  
 QY 320 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLysPro 339  
 |||||.....  
 Db 2060 CACACACCTGTGCACTGGAGAGAACACCCCATGTGCTCTTTGTTCTATGAGGAGGCCA 2119  
 QY 340 Gln----- 340  
 Db 2120 CAGGTGACAGTGTGTAACAACCTCTTAAAGGAGCATCAATACCTCGGCGTAAATACACTG 2179  
 QY 340 ----- 340  
 Db 2180 GCATCTTGGGCTATGCTGTGTGTGATGATGATGTGGGCTCTGTGTACGCGGCGCTTG 2239  
 QY 340 ----- 340  
 Db 2240 CACTTGAGAGGCGGCGCTGAATAATCAATATGGCCAGGTGAGATTGAGGACCAAGGTGAA 2299

QY 340 ----- 340  
 Db 2300 GGCTTGACATAGGTGGCTGAGAAATATGAGCTTCATGATGAGCGGAGTGCATCAT 2359  
 QY 340 ----- 340  
 Db 2360 GGCTGTCTTACGGCGCTTCTCTCATCATGAGGCTCATCCAGGCCAAGGTGTC 2419  
 QY 341 ---ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 359  
 |||||.....  
 Db 2420 AAGCTTAGCTTCCGGGCGCTCTCTGCTCATCTGTGTGATGAGCTTATACACAGGTTACAG 2479  
 QY 360 GluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 379  
 |||||.....  
 Db 2480 GAACGATACATGATGATCTCCCGCAAAATATCAAGCAAGGCTATGAGCAGAGGTGTGTAGCC 2539  
 QY 380 MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeu 399  
 |||||.....  
 Db 2540 CTGCATGTGAGAGAGTGCAGATGAGCCTTAACCGCTCTTATCTCCACAGGCTTCTG 2599  
 QY 400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419  
 |||||.....  
 Db 2600 GACGAAAGCTTCACTTCTTCCACACAAATTTCTGTGTGCTCCAGCTGATCCAGACAGA 2659  
 QY 420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439  
 |||||.....  
 Db 2660 AAGCATACAGCTTCAATCATCAATCCCAAGAGAGACATAGCATCCGCTGCGCGAGTCC 2719  
 QY 440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrLysGluGluAsnLeu 455  
 |||||.....  
 Db 2720 GGAGACATATTACAGAGGTGAGCTGTGCTGCTTCTGACAGAACACCTG 2767  
 RESULT 6  
 BX372276/c  
 LOCUS BX372276 910 bp mRNA linear EST 08-MAY-2003  
 DEFINITION Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 ACCESSION BX372276  
 VERSION BX372276.1 GI:30448117  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1. W.B. Gruber, C., Jesse, J. and Polyes, D. Full-length cDNA libraries and normalization Unpublished  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7542.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BA1043ZH06.CS04080.1ccluster=7542.r.  
 Contact: Reng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID: CS0BA1043ZH06\_CS04080\_1.  
 Location/Qualifiers  
 1. 910  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL005YD02"  
 /cell\_line="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /clone\_lib="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V

BASE COUNT sites of the PCMVSPORT 6 vector. Library was normalized.  
 251 a 193 c 176 g 288 t 2 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,24e-143 Length: 910  
 Score: 1397.00 Matches: 282  
 Percent Similarity: 96.91% Conservative: 0  
 Best Local Similarity: 96.91% Mismatches: 9  
 Query Match: 56.29% Indels: 4  
 DB: 13 Gaps: 0

US-10-070-464-5 (1-465) x BX372276 (1-910)

1 ThrGlyThAlaAsnProlyValThrPheMetSerGluLeuMetIleAspAlaGlu 20  
 869 ACAGGACAGCAAACTCTTAAGTCACTTTTAAAGTCAAAATAATGATGATGCTGAA 810  
 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleAspGlu 40  
 809 GAAGGATCATATGATCATAGTAAAGAACTAATTCACCTTTGAGATTCATT-GAA 752  
 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 751 GNAAGTGGATATATGCGAGAGCTGATGAGTCACTCTGAGGAAATATATGCTGCATC 692  
 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 691 CTAATGATGATGCTCCAGACTCCCTCAAAATAGTGTGATCCACTGATTAATTTATTC 632  
 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 631 NCAGTGAAGATATATGATGAAAGGACAGACTCATTTGACATGCTGATTCGTG 572  
 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 571 AGCCCACTAATTCATATGAGAAACAAACAGACATCGATAATATCCATGACATCTTT 512  
 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAspSerGluCysGly 140  
 511 CATGTTTTTCCCAAGTCAAGAAAGAAATGAGTTATTTTCTCTGATGCACAA 452  
 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 451 AAGAGTTCCGTCATTTATACAAATTCATCTAATTTAAAGAAAGCAAAATTAACGA 392  
 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 391 TCCAGTGGTGGCTGCTGCTCCAAAGTATTCAGTGTCTATCAAAAGAGAGATAGCA 332  
 181 ILeThrSerGlyGluTyrProGluValIleGluYargHisGlySerAsnIleGlnValAspGlu 200  
 331 ATTACAGAGTGAATGGAGAGTCTTGCGCGGACGAGATCAATATCCAAAGTTGATGAA 272  
 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLysTyr 220  
 271 GTCAGAAAGCTGTATATTTTGAAGGACCAAAAGATCCCTTTAAGCATCACTGTAC 212  
 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 211 GATAGTACGTTACGTAATCTCGAGAGGTGACAAAGCTGACTGACCGTGGCTACTACAT 152  
 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsn 260  
 151 TCTTGCTGATCAGTACAGACTGTGACTCTTTATATAGTAAGTATAGTAAACCAAGAAT 92  
 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 91 CCAACTGTGTCTCTTTTCAAGCTATCAAGTCTCGAAGATGATCCCAAC-TCCAAACA 33  
 281 LysGluPheTyrPalaThrIleLeuAspSerAla 291  
 32 AAGAAATTTGGCGCACCAT-TTGATTCAGCA 1

RESULT 7  
 BI223892  
 LOCUS  
 DEFINITION  
 602941035F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:5104113 5',  
 mRNA sequence.  
 BI223892  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 853)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHAM1251 row: e column: 10  
 High quality sequence stop: 718.  
 Location/Qualifiers  
 1..853  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5104113"  
 /cruise\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_12"  
 /note="Organ: cervix; Vector: pCMV-SPORT6, Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."

BASE COUNT 250 a 175 c 187 g 241 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4e-129 Length: 853  
 Score: 1265.50 Matches: 235  
 Percent Similarity: 96.36% Conservative: 3  
 Best Local Similarity: 95.14% Mismatches: 6  
 Query Match: 50.99% Indels: 1  
 DB: 12 Gaps: 1  
 US-10-070-464-5 (1-465) x BI223892 (1-853)  
 108 GluThrThAspIleTyrPheIleAsnIleHisAspIlePheHis-ValPheProGlnSerHis 127  
 61 GAAGAAACAGACATTCGATTAATATTCATGACATCTTTCAGTGTTCCTCCCAAGTCA 60  
 127 sgluglugluileglupheilepheleaserglucyslythrnglyphearghisleuty 147  
 61 CGAAGAGGAAATGAGTTATTTTGGCTGATGATCAAAAGATTCCTCGTCATTTATTA 120  
 147 rlysllethrserleleuilegluseryslytyrlysaraserseglyyleuproal 167  
 121 CAAATATCATCTATTTTAAAGAAAGCAAAATATAACATCCAGTGGTGGCTGCTGC 180  
 167 aProserAspPheLysCysProIleLysGluGluIleAlaIleThrSerGlyGluTyrPro 187  
 181 TCCAAAGTATTTCAAGTGTCTATCAAAAGAGAGATGAGAAATTTACAGTGGTGAATGGA 240  
 187 uValleuGlyArgHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrPh 207  
 241 AGTCTTGCCCGCATGATCTAATATTCAGATGATGAAGTCAAGAGCTGTATATTT 300

QY 207 eGIUGlYThrIysAspSerProLeuGluHisIleuTYrValIserTYrValAsnPr 227  
 DB 301 TGAAGGACCAAGAGCTCCCTTTAGAGCATCAGCTGATAGCTTACGTAAATCC 360  
 QY 227 OGIVGluValThrArgLeuThrAspArgLYrYrSerHisSerCysGlyIleSerGlnH 247  
 DB 361 TGGAGAGGTGACCAAGCTGACTGACCGGTGACTACCATCTTGCTGCATCAGTCACAA 420  
 QY 247 sCyAspPhePheIleSerLYrYrSerAsnGlnIlyAsnProHisCysValIserLeuTY 267  
 DB 421 CTGGACTCTCTTTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGT 480  
 QY 267 rLYsLeuSerSerProGluAspAspProThrCysLYsThrLYsGluPheTrpAlaThrI 287  
 DB 481 CAACTATCAAGTCTCTGAAAGATGACCCAACTTGCAAAACAAAGAAATTTGGCCACCAT 540  
 QY 287 eLeuAspSerLYsArgProLeuProAspTYrThrProProGluIlePheSerPheGlu 307  
 DB 541 TGTGGATTTCAGCAGCTCTCTCTCTGACTATATCTCTCAGAAATTTCTCTTTGAAAG 600  
 QY 307 rThrThGlyPheThrLeuTYrGlyMeLeuTYrLYsProHisAspLeuGlnProGlyLY 327  
 DB 601 TACTACTGAGATTTCATTTGATGGATGCTCTACAGCCTCATGATCTACAGCCTGAAA 660  
 QY 327 sLYsTYrProThrValLeuPheIleTYrGlyLYrProGlnVal-----Al 342  
 DB 661 GAAATATCTCTACTGCTGCTTCATATATGCTGCTCAGAGTGCAGTGTGATTAATCG 720  
 QY 342 aIleAlaGlyAlaProVal 348  
 DB 721 AGTTAAAGAGATCCAGTA 739  
 RESULT 8  
 BG259714 993 bp mRNA linear EST 13-FEB-2001  
 LOCUS 60238007F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4511080 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG259714  
 VERSION BG259714.1 GI:12769530  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 993)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: gsp@bbs-rcmail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LLM10393 row: 0 column: 17  
 High quality sequence start: 6  
 High quality sequence stop: 656.  
 Location/Qualifiers  
 1..993  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4511080"  
 /issue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC\_92"  
 /note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: Not1;  
 Site\_2: Sal1; Cloned unidirectionally; Oligo-dT primed.  
 Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT	296 a	183 c	243 g	271 t
ORIGIN				

Alignment Scores:  
 Pred. No.: 5,41e-122 Length: 993  
 Score: 1202.00 Matches: 232  
 Percent Similarity: 90.42% Conservative: 4  
 Best Local Similarity: 88.89% Mismatches: 11  
 Query Match: 48.43% Indels: 14  
 DB: 10 Gaps: 5

US-10-070-464-5 (1-465) x BG259714 (1-993)

QY	93	IIleGluSerValProAspSerValThrProLeuIleIleTYrGluGluThrThrAspIle	112
DB	2	ATTGAGTCAGTGCCTGATTCGTGACCCACTAATTTATCTATGAGAAACAAACAGACATC	61
QY	113	TrpIleAsnIleHisAspIlePheHisValPheProGlnSerHisGluGluGluIleGlu	132
DB	62	TGGATTAATATCCATGACATCTTTCAATGTTTTCCCAAAGTCAGAAAGAAATTCAG	121
QY	133	PheIlePheAlaSerGluCysLYsThrGlyPheArgHisIleuTYrLYsIleThrSerIle	152
DB	122	TTTATTTTTCCTCGAATGCAAAACAGTTTCCGTCATTTATCAAAATTAATCAATCTATT	181
QY	153	LeuLYsGluSerLYsTYrLYsArgSerSerGlyLYsLeuProAlaProSerAspPheLYs	172
DB	182	TTTAAAGAAAGCAAAATTAACGATCCAGTGTGTGGCTGCTCCAAAGTATTAATTCAG	241
QY	173	CysProIleLYsGluGluIleAlaIleThrSerGlyGluTrpGluValLeuGlyArgHis	192
DB	242	TGTCTTATCAAAAGAGAGATAGCAATTAACAGTGTGTAAATGGAAATTTGGCCGGCAT	301
QY	193	GlySerSerIleGlnIleValArgGluValArgArgLeuValTYrPheGluGlyThrLYsAsp	212
DB	302	GGATCTAATATCCAAAGTTGATGAAGTCAGAAAGCTGTGATAATTTTGAAGGACCAAAAGAC	361
QY	213	SerProLeuGluHisIleuTYrValIleValSerTYrValAsnProGlyGluValThrArg	232
DB	362	TCCCTTTAGAGCATCCTGTAGCTGATGATGATTAATCTGGAAGAGTACAAAG	421
QY	233	LeuThrAspArgGlyTYrSerHisSerCysLYsSerGlnHisCysAspPhePheIle	252
DB	422	CTGACTGACCGTGGCTACTCACAATTTCTGCTGATCAGTCAGCATCTGACTTTTATA	481
QY	253	SerLYsTYrSerHisGlnIlyAsnProHisCysValSerLeuTYrLYsLeuSerSerPro	272
DB	482	AGTAAAGTATGTAAACCAAGAAATCCACACTGTGTGCTTTTACCAAGCTATCAAGTCCT	541
QY	273	GluAspAspProThrCysLYsThrLYsGluPheTrpAlaThrIleLeuAspSerAlaGly	292
DB	542	GAAATATACCACTTCACAAACAAAGAAATTTGGGCCACCATTTGGATTCAGCGGT	601
QY	293	ProLeuProAspTYrThrProProGluIlePheSerPheGluSerThrTrnGlyPheThr	312
DB	602	CTCTTCCTGACTATACCTCTCCGAAATTTTCTTTGAAAGTACTACGATTTACC	661
QY	313	Leu---TYrGlyMeLeuTYrLYsProHisAspLeuGlnProGlyLYsLYsTYrProThr	331
DB	662	CTTGTGTGGGGAGT-----CCTTACCAAG-----	688
QY	332	ValLeuPheIleTYr---GlyGlyProGlnValAlaIleAlaGlyAlaProValThrLeu	350
DB	689	GCCTTATATGATCTACAAAGCTGGAGAGAAATTTCTTAATGGCGT---GTCAATTAT	745
QY	351	Trp 351	
DB	746	TGG 748	
RESULT 9	0	*	
BG974587			



Tissue Procurement: T. Rajendra Kumar and Martin M. Matzuk  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 plate: NDCM08 row: b column: 19  
 High quality sequence stop: 551.  
 Location/Qualifiers

## FEATURES

source

```

1.774
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30248202"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_166"
/notes="Organ: brain; Vector: pDNR-LIB; Site 1: SfiI
(ggcatatggcc); Site 2: SfiI (ggcgcctggcc);
Non-normalized full-length enriched library 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTAAGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 2.05
kb (range 1.0-4.0 kb). 15/15 clones contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Corp. "
```

BASE COUNT 228 a 187 c 160 g 198 t 1 others

ORIGIN

## Alignment Scores:

Pred. No.: 6,41e-114 Length: 774  
 Score: 1127.50 Matches: 221  
 Percent Similarity: 89.88% Conservative: 10  
 Best Local Similarity: 85.99% Mismatches: 19  
 Query Match: 45.43% Indels: 7  
 DB: 14 Gaps: 2

US-10-070-464-5 (1-465) x CB233750 (1-774)

```

QY 51 ThProgluGlySerTyraATPSeTleLeuLeuAspArgSerGlnThrArgLeuGln 70
DB 6 ACTCCAGAGGAGAAACATCGCTGCTCATCTACTACACCGTCCCGACTGACCTGCGAG 65
QY 71 lLeValLeuLeuSerProgluLeuPheLeuProValGluAspValMetGluArgGln 90
DB 66 ATAGTTCTGATCTCCCTGAGTTATTCATCCAGTAGAAGATGCGACAGACAG 125
QY 91 ArgLeuileGluSerValProAspSerValThrProLeuileTleTyrgluGluThr 110
DB 126 AGACTTATAGAGTCAGTCTGACTGTGACACCACTGATCATGTGAAGAAACACA 185
QY 111 AspIleTrrPLeaenIleHisAspIlePheHisValPheProGlnSerHisgluGlu 130
DB 186 GACATCTGGATTAATATCCACGATATTTTCATGATTTTCTCMAACTCATGAAGTAA 245
QY 131 lLeGluPheIlePheAlaSerGluCysLeuThrGlyPheArgHisLeuTyrlsIleThr 150
DB 246 ATTGAGTTATTTTTCCTCTGTAATGCAAAACAGTTTCTCATCTGTATTAATACACA 305
QY 151 SerIleLeuLeuSerIleTyrlsArgSerSerGlyGlyLeuProAlaProSerAsp 170
DB 306 TCATTTTAAAGAGGACAAATATTAACGGTCCAGTGTGAGTACTGCCCCAAGGAT 365
QY 171 PheLeuCySerProIleLeuSerGluIleAlaIleHisSerGlyGluTrrPgluValLeuGly 190
DB 366 TTCAAGTGTCCTTCAAAAGAAATTAACAATTAACGCTGTAATGGAAGTAACTTGCG 425
QY 191 ArgHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrrPhegluGlyThr 210
DB 426 CGGCAATGATCTTAATATCTGGGTTGATGAAGCAAGAAAGCTGCTGTTCTTGAAGGACC 485
```

```

QY 211 LyAspSerProLeuGluHisHisLeuTyrlsValSerTyrlsValAsnProgluGluVal 230
DB 486 AAGAGCTCTCTTTGGAGCATCATCTGTAGTACCAAGTATATCAAACTGGAGAAAGTG 545
QY 231 ThrArgLeuThrAspArgGlyTyrlsSerHisSerCysIleSerGlnHisCysAspPhe 250
DB 546 GTGAGGCTGACTGACCGGTGGCTACTCACACTCTGCTGCTGACGCGGCAATGGATTC 605
QY 251 PheIleSerTyrlsSerLeuGlnIleLysAsnProHisCysValSerLeuTyrlsLeu 270
DB 606 TTCATTAAGTAGTAGACCAACCAAGAAATTCACATGTGTCTTCCCTTACCAACTCT 665
QY 270 erSerProgluAspAspProThrCys---LysThrLysGluPheTrrPalaThrIleLeu 288
DB 666 CAAGTCCCTGAGATACCACTGATTTGATAAAACAAAGAAATTTGGCCACCAATTT 725
QY 289 -----AspSerAlaClyProLeuProAspTyrlsThrProPro 300
DB 726 TGGATTCACCAAGCAAGGTCCTCTTCTGTAAGTACCAACCCCT 768
```

## RESULT 11

BM557438

LOCUS

DEFINITION

BM557438 1042 bp mRNA linear EST 20-FEB-2002  
 AGENCOURT\_6578992 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5466759

5', mRNA sequence.

ACCESSION

BM557438

VERSION

BM557438.1 GI:18799430

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

NIH-MGC <http://mgs.nci.nih.gov/>.  
 1 (bases 1 to 1042)

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph. D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

TITLE

JOURNAL

COMMENT

CDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 plate: LLCM1967 row: k column: 16  
 High quality sequence stop: 697.  
 Location/Qualifiers

## FEATURES

source

```

1.1042
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5466759"
/clone_1lb="NIH_MGC_41"
/notes="Organ: skin; Vector: pOT87; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGACAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

BASE COUNT 279 a 223 c 234 g 303 t 3 others

ORIGIN

## Alignment Scores:

Pred. No.: 1.31e-112 Length: 1042  
 Score: 1117.50 Matches: 225  
 Percent Similarity: 65.52% Conservative: 3  
 Best Local Similarity: 64.66% Mismatches: 14  
 Query Match: 45.02% Indels: 106



```

DB: 12 Gaps: 2
US-10-070-464-5 (1-465) x BM557438 (1-1042)

QY 187 GluValIleuGlyYArgHISerAenIIeGlnValAAspGluValAArgAglLeuValTyr 206
DB 1 GAAGTCTTGCGCCGCGATGATTAATATCCAAAGTTGATGATGATGATGATGATGATGAT 60
QY 207 PheGluGlyThrIleuAAspSerProIleuGluHISerIleuTyrValAAspTyrValAAsn 226
DB 61 TTTGAAGGACCAAGACCTCCCTTTAGACATACATGATGATGATGATGATGATGATGAT 120
QY 227 ProGluValIleuThrArgLeuThrAAspArgGlyTyrSerHISerCysCysIleSerGln 246
DB 121 CTGAGAGAGGTGACAAAGGCTGATGACCGGTGCTACTACATCTTGCTGATCAGCAG 180
QY 247 HISerAAspPheHISerIleuTyrSerAAsnGlnIleAAsnProHISerCysValSerLeu 266
DB 181 CACTGTGACTCTCTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 240
QY 267 TyrIleuSerSerProGluAAspAAspProThrCysIleuThrIleuGluPheTyrValAThr 286
DB 241 TACAAAGCTATCAAGGCTGTAAGATGACCAACTTGAACAAAGAAATTTGGGCGCACC 300
QY 287 IleuAAspSerAAspIleuProIleuProAAspTyrThrProProGluIlePheSerPheGlu 306
DB 301 ATTTGGATTCAGCAGGCTCTCTCTGCTACTATCTCTCCAGAAATTTCTCTTTGAA 360
QY 307 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleuProHISerAAspLeuGlnProGly 326
DB 361 AGTACTAGCTGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
QY 327 LysIleuTyrProThrValIleuPheIleuTyrGlyGlyProGln----- 340
DB 421 AAGAAATATCTCACTGCTGCTCATATATATGATGATGATGATGATGATGATGATGAT 480
QY 340 ----- 340
DB 481 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTAGTATATGATGTT 540
QY 340 ----- 340
DB 541 GTATGATATAGACAAAGGAGATCTGTCAACGAGGCTTAAATTGAAAGCGCTTTANA 600
QY 340 ----- 340
DB 601 TATAAATGGGTCAAAATAGAAATTTGACGATCAGTGGAGGAGACTCCAAATATCTAGCTCT 660
QY 340 ----- 340
DB 661 CGATATGATTTCAATTGACTTACATGCTGTGGGATCCAGGCTGCTCTATGAGAGATAC 720
QY 341 -----ValAAspIleuAAspIleuAAspIleuAAspIleuAAspIleuAAsp 346
DB 721 CTCTCCCTGATGATTAATGACAGAGTCAAGATATCTTCAAGGTTCTATGCTGGGCGC 780
QY 347 ProValIleuThrIleuPheTyrAAspThrGlyTyrThrGluArgTyrMetGlyHISerPro 366
DB 781 CCAAGTACTCTGTGGATCTTCTATGATACAGATACAGGATGATGATGATGATGATGAT 840
QY 367 AAspGlnAAspGlnIleuTyrTyrIleuGlySerValAAspGlnAAspIleuAAspPhePro 386
DB 841 GACCAAGATGAAAGAGGCTATTAATCTATGATGATGATGATGATGATGATGATGATGAT 900
QY 387 SerGluProAAspAAspIleuLeuLeuHISerGlyPheLeuAAspGluAAspValHISerAAsp 406
DB 901 TCTGAACCAAAATGATTAATCTGCTTACATGCTTTTCTGATGATGATGATGATGATGAT 960
QY 406 IAspIleuThrSerIleuLeuSerPheLeuValAAspIleuAAspIleuAAspIleuAAsp 424
DB 961 CACATATCCAGATATTAATCTGAGGTTTCTTTTGTGAGGAGGCTGAAAGAGCCCATATGAT 1020
QY 425 -----GlnIleuTyrPro 428

```

```

DB 1021 TTTACAGCAATCTATCTCT 1038
RESULT 12
BI084090 1041 bp mRNA linear EST 20-JUN-2001
LOCUS 602869453F1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5013996 5',
DEFINITION mRNA sequence.
ACCESSION BI084090
VERSION BI084090.1 GI:14502420
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukavota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1041)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Invitrogen, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LLCM1820 row: j column: 13
High quality sequence stop: 847.
Location/Qualifiers
1. 1041
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5013996"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_102"
/note="Organ: salivary gland; Vector: pOT87; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 280 a 227 c 242 g 292 t
ORIGIN
Alignment Scores:
Pred. No.: 1.11e-108 Length: 1041
Score: 1082.00 Matches: 217
Percent Similarity: 63.43% Conservative: 5
Best Local Similarity: 62.00% Mismatches: 15
Query Match: 43.59% Indels: 113
DB: 12 Gaps: 4
US-10-070-464-5 (1-465) x BI084090 (1-1041)
QY 157 LysTyrIleuAAspSerSerGlyIleuProAlaProSerAAspPheIleuSerProIleu 176
DB 3 AAATATTAACGATCCAGTGGGCTGCTGCTCCCAAGTATTAAGTCTCTATCAAA 62
QY 177 GlnGluIleuAlaIleuThrSerGlyIleuTyrGluValIleuGlyYArgHISerAAspIleu 196
DB 63 GAGGAGATGACATTATCCAGTGGTGAATGGAAGTCTTGCGCGATGATGATTAATATC 122
QY 197 GlnValAAspGluValAArgAglLeuValTyrPheGluGlyThrIleuAAspSerProLeuGlu 216
DB 123 CAAGTTGATGAAGTCAAGAGGCTGGTATATTTTGAAGGACCAAAAGCTCCCTTTAGAG 182
QY 217 HISerIleuTyrValIleuSerTyrValAAspProGluValIleuThrArgLeuThrAAspArg 236

```

Db	183	CATCACCTGTAAGTAGTACGTACCTAATCTCGAAGAGTGACAGGCTACCTGACCGCT	242
Oy	237	GLYTYRSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSer	256
Db	243	GGTACTACACATTCCTTGCTGCATCAGTCAGACACTGTGACTCTTTATGAAGATATAGT	302
Oy	257	AsnGlnIysAsnProHisCysValSerLeuTyrIleValSerSerProGlnAspAspPro	276
Db	303	AACCAAGAAATCCACACTGTGTGTCCCTTACAAAGCTATCAAGTCTCGAAAGATGACCCA	362
Oy	277	ThrCysIleThrIleGluPheTyrPheIleThrIleLeuAspSerAlaGlyProLeuProAsp	296
Db	363	ACTTGCAAAACAAAGAAATTTTGGGCCACCAATTTGGATTGATCAGACGGTCCCTTCCTGAC	422
Oy	297	TyrThrProProGlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet	316
Db	423	TATATCTCTCCAGAAATTTCTCTTTAAAGTACTACTCGAATTTACATTTATGCGGATG	482
Oy	317	LeuTyrIleProHisAspLeuGlnProGlyIleValSerTyrProThrValLeuPheIleTyr	336
Db	483	CTTACAAAGCCCTCAATGATCTACAGCCTCGAAAGAAATATCTACTGTGCTCTTCATATAT	542
Oy	337	GlyGlyProGlnValAlaIleAla-----	344
Db	543	GGTGTCTCTCCAGTGTCAGTGTGTGAATATGCTTTAAACAGCAGTCAATATTTCCGCT	602
Oy	344	-----	344
Db	603	TGAATACCTTAGCCTCTAGTTATGTGGTTGTAGTATGATAGACAAACAGGGATCTGTCTC	662
Oy	344	-----	344
Db	663	ACCGAGGCGTTAAATTTGAAGCGCCTTAAATATACAAATGGGTCAATGAAATTTGACGAT	722
Oy	344	-----	344
Db	723	CACGTGAGAGACCTCCAAATTATCTAGCTTCTCGATATGATATTTCAATTGACTTAGATCGTG	782
Oy	344	-----	344
Db	783	TGGGCATCCACGGGTGTCTTATGAGAGGATACCTCTCCCTGAATGGCATTATGACAG	842
Oy	345	-----GlyAlaProValThrLeuTrpIlePhe-Tyr	354
Db	843	GTCAAGATACCTTCCAGGGGTGCATATAGCTGGGGGCCACAGTCACCTGTGATCTTCTTA	902
Oy	354	ASPThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTy	374
Db	903	TGATACAGAGATACACGAAACGTTATATGGGTACCTGAAACAGA---TGAACGGGTCAATC	959
Oy	374	rLeuGlySerValAlaMetGlnAlaGluIlePhePro-----Se	387
Db	960	TTTCAGGATCTGTG-----CCTGCAGCGCAAAATTTCCCTTGACAAATCGTTAATGCTCAAT	1013
Oy	387	rGluProAsnArgLeuLeuLeuHis	396
Db	1014	GGTTCGGGCAAAATTCCTTGATACCAAC	1041
RESULT 13			
LOCUS	CD352521	762 bp	mRNA linear EST 29-MAY-2003
DEFINITION	UI-M-G10-CGK-e-23-0-UI_r1 NIH BMP_G10	Mus musculus	CDNA clone
ACCESSION	CD352521		IMAGE: 6856416 5', mRNA sequence.
VERSION	CD352521.1	GI:31144032	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	1 (bases 1 to 762)		
	NIH-MGC	http://mgi.nci.nih.gov/.	

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@osf-mail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Distribution information can be found at <a href="http://genome.uiowa.edu/distribution/mouse1.html">http://genome.uiowa.edu/distribution/mouse1.html</a> This clone was contributed by the Brain Molecular Anatomy Project of BMAP.
FEATURES	Seq primer: pyx-5.
SOURCE	Location/Qualifiers 1..762 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE: 6856416" /tissue_type="whole brain" /dev_stage="embryo 13.5,14.5,16.5,17,5dpc" /lab_host="DH10B (T1 phage resistant)" /clone_1nb="NH_BMAP_G10" /note="Organ: Brain; Vector: pyx-Asc; Site_1: Ecor I; Site_2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACG. This library was created for the polya tail Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT	228 a 158 c 169 g 205 t 2 others
ORIGIN	
Alignment Scores:	
Pred. No.:	8,59e-107 Length: 762
Score:	1063.00 Matches: 205
Percent Similarity:	95.09% Conservative: 8
Best Local Similarity:	91.52% Mismatches: 11
Query Match:	42,83% Indels: 1
DB:	14 Gaps: 0
US-10-070-463-5 (1-765) x CD352521 (1-762)	
Gy	1 ThrG1ThralaaanProLyseValThrPhelysmetSerGiulleMelelaaspAlaglu 20
Ddb	82 ACAGGACAGCAAAACCAGGTCACTTTCAAGATGTCGGAGATTGTTGTAAGCTGCCA 14
Gy	21 GlYArgllellaespyallleaaplysgluleulleginProphegluullealeuPheglu 40
Ddb	142 GGAGGGATTATATATATCATATAGATAAGAACGGTTCAACTTGCGAATTCCTTTGAG 201
Gy	41 GlVValGlutyrillelaaragaiaGLYTPTRrProgiugLiLYSTYALATrSrile 60
Ddb	202 GGAGTTGAATATTTCCAGACCGGAGTGCATCCAGAGGAAAAACAAGCTGGTCATC 26
Gy	61 LeuluaspaRserginThraRgleugnilleValleulleleserProgiuleuPhele 80
Ddb	262 CTACTGACCGTTCCAGACTCACCTGCAGAGAAGTTGATCTCCCCTGAGTTATTCATC 321
Gy	81 ProValGlunaspasyalMerGluarrginargleulleguiserValProaspSerVal 100
Ddb	322 CAGTGAAGATATATCCATGGACACACAGAACTTATAGAGTCAGTTCGACTCTGTG 381
Gy	101 ThnProleullelleTyrlgLuigluthThraspietripileanuilehsaspillephe 120

```

Db      382  ACACCATGATCATCTATGAGAAACACACACATCTGGATTAATATCATCAGATATTTT 441
Qy      121  HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIys 140
Db      442  CATCTTTTCTCTCAAACTCATGAAGATGAATGATGATTTTCTCTGTAATGCAAA 501
Qy      141  ThrGlyPheArgHisIleuTyrlsIleThrSerIleuIysGluSerIleTyrlsArg 160
Db      502  ACAGCTTTTCTCTCATCTGTATTAATATCATCTTTTAAAGAGACAAATATTAACGG 561
Qy      161  SerSerGlyIleuProIleProSerAspPheIysCysProIleIleGluIleAla 180
Db      562  TCCAGTGTGGAGTACCTGCCCCCAAGTATTCAGTGTCTCTATCAAAAGAAATATA 621
Qy      181  IleThrSerGlyIleuTyrGluValLeuGluIleArgHisGlySerAsnIleGluValAspGlu 200
Db      622  ATTACCGATGTGTGAATGGAGAGTACTGGCCGCGCATGATCTAA-ATCTGGTTGATGAA 680
Qy      201  ValArgArgIleuValIleTyrlsPheGluGlyThrIysAspSerProLeuGluIleHisIleuTyrls 220
Db      681  GCCGAAAGCTGTGCTACTTTGAAAGCCACCAAGACTCTCTTTGAGCATCACCCTGAC 740
Qy      221  ValValSerTyrls 224
Db      741  GTGACCAAGTAT 752

```

```

RESULT 14
AL040398 753 bp mRNA linear EST 29-FEB-2000
LOCUS DKEZp434A0714_r1.434 (synonym: hce3) Homo sapiens cDNA clone
DEFINITION DKEZp434A0714_5', mRNA sequence.
ACCESSION AL040398
VERSION AL040398.1 GI:5409350
SOURCE EST.
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 753)
AUTHORS Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE EST (Koehler, et al.)
JOURNAL Unpublished
COMMENT Contact: Koehler K
MIPS
Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMPZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
No st sequence available.
This clone (DKEZp434A0714) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

```

## FEATURES

```

source
1..753
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKEZp434A0714"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="434 (synonym: hce3)"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"
BASE COUNT 218 a 150 c 170 g 214 t 1 others
ORIGIN

```

```

Alignment Scores: 6.47e-106 Length: 753
Pred. No.: 1055.00 Matches: 193

```

```

Percent Similarity: 99.49% Conservatve: 1
Best Local Similarity: 98.97% Mismatches: 1
Query Match: 42.51% Indels: 0
DB: 9 Gaps: 0
US-10-070-464-5 (1-465) x AL040398 (1-753)

```

```

Qy      149  IleThrSerIleuIysGluSerIleTyrlsArgSerGlyIleuProAlaPro 168
Db      1  ATTACATCTTATTTTAAAGAAAGCAAAATATTAAGATTCAGTGTGGCTCTGCTTCA 60
Qy      169  SerAspPheIysCysProIleIleGluIleAlaIleThrSerGlyIleuTyrGluVal 188
Db      61  AGTATTTTCAAGTGTCTCTATTAAGAGAGATAGCAATTAACAGTGTGATATGGAAATT 120
Qy      189  LeuGluIleArgHisGlySerAsnIleGluValAspGluValArgArgIleuValIlePheGlu 208
Db      121  CTTGGCCGCGCATGATCTAATATCAAGATTGATGATGATGATGATGATGATGATGATGAT 180
Qy      209  GlyThrIysAspSerProLeuGluIleHisIleuTyrlsValIleSerTyrlsValAspProGly 228
Db      181  GGCACCAAGACTCCCTTTAGACATCACCTGTACGTACGTACGTACGTACGTACGTACGTAC 240
Qy      229  GluValThrArgIleuThrAspArgGlyTyrlsSerHisSerCysIleSerGluHisCys 248
Db      241  GAGGTGACAAAGCTGACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Qy      249  AspPhePheIleSerIleTyrlsSerAsnGluIleAspProHisCysValSerIleTyrls 268
Db      301  GACTCTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 360
Qy      269  LeuSerSerProGluIleAspAspProThrCysIleTyrlsGluPheTyrPalaThrIleLeu 288
Db      361  CTATCAAGCTCCTGAAGTGAAGCCCAACTTGCAAAAGAAATTTGGCCACCATTTTG 420
Qy      289  AspSerAlaGlyProLeuProAspTyrlsThrProProGluIlePheSerPheGluSerThr 308
Db      421  GATTACAGAGTCTCTCTCTGACTATACCTCTCAAGAAATTTCTTTTGAAGACT 480
Qy      309  ThrGlyPheThrIleuTyrlsMetLeuTyrlsPheProHisAspLeuGluProGlyIleIys 328
Db      481  ACTGATTTTACATTTGATGGAGAGCTCTACAGCTCATGATCTACAGCCCTGGAAGAA 540
Qy      329  TyrProThrValIleuPheIleTyrlsGlyProGluValAlaIle 343
Db      541  TATCTTACTGTGCTGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 585

```

```

RESULT 15
BG974324 831 bp mRNA linear EST 12-JUN-2001
LOCUS BG974324
DEFINITION 602844075F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979707 5',
mRNA sequence.
ACCESSION BG974324
VERSION BG974324
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 831)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: CGAPs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

Plate: LHAM10978 row: e column: 20  
High quality sequence stop: 797.  
Location/Qualifiers  
1. .831  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4979707"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_id="NCL\_CGAP\_Mam4"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."

BASE COUNT 235 a 191 c 190 g 215 t  
ORIGIN

Alignment Scores:  
Pred. NO.: 3.46e-104 Length: 831  
Score: 1040.00 Matches: 190  
Percent Similarity: 93.75% Conservative: 5  
Best Local Similarity: 91.35% Mismatches: 13  
Query Match: 41.90% Indels: 0  
DB: 12 Gaps: 0

US-10-070-464-5 (1-465) x BG974324 (1-831)

QY 136 AlAserGIuCySLySThGlyPheARGHISLeuTYrLYSLerhSerILEuLYSGlu 155  
|||  
|||  
6 GCGTCCGGGAGGCGGTGGGTTTGTGTCATCTGTATAAATCACAATCATTTTAAAGGAG 65  
QY 156 SerLYSTYrLYSArGSeTSeTGLyLYeUProAlaProSeRaSPheLYCySProlle 175  
|||  
66 AGCAAAATATAAACGGTCAGTGAGTGAGTACCTGCCCAAGTGATTTCAAGTGTCTTATC 125  
QY 176 LysGIuGIuILEAlleThrSeGIyGIuTrPGIuValleuGLyARGHISGLYSerAsn 195  
|||  
|||  
126 AAAGAGAAATTAACAATTACCAAGTGTGATGGAGAGTACTTGGCCGCAATGATCTTAT 185  
QY 196 IleGIuValAspGIuValArgArGLeuValTYrPheGIuGIyThrLYSASPSeProleu 215  
|||  
|||  
186 ATCTGGGTTGATGAAGCAGAAAGCTGTGTACTTTGAAGGACCAAAAGACTCTCTTG 245  
QY 216 GluHISHisLeuTYrValValSerTYrValAsnProGIyGIuValThrArgLeuThraSP 235  
|||  
246 GAGCATCACTGTGTGATGACCAAGTATGCAAACTCTGAGAAAGTGAGGCTGACTGAC 305  
QY 236 ArgGIYrYrSerHISerCySeSeriLeSeGIuHISCySASPhePheILESerLYSTYr 255  
|||  
|||  
306 CGTGCTTCTACACTCTGCTGCTGCTGACCGGCACTTGTGACTTCTTCAATAGTAAGTAC 365  
QY 256 SerASnGIuLYSAsnProHISCySValSerLeuTYrLYSLeuSerSeProGIuAspASP 275  
|||  
|||  
366 AGCAACCGAAGAAATCCACACTGTGTGTCCTCTACAAACTCTCAAGTCTGAGAGATGAC 425  
QY 276 ProThrCySLySThLYSGluPheTrPAIaThrILEuAspSerAlaGLYProleuPro 295  
|||  
|||  
426 CCAAGTTCAATAAACAAAGAAATTTGGGCCACCATGTGATTCAGCAGGTCCTCTTCT 485  
QY 296 AspTYrThrProPGIuILEPheSerPheGIuSerThrThrGLYPhetThrLeuTYrGLY 315  
|||  
|||  
486 GACTACACCCCTCCAGAAATTTTCTTTTGAAGTACTACTGATTTACACTGTATGGA 545  
QY 316 MetLeuTYrLYSProHISAspLeuGIuProGLYLYSLYSTYrProThrValleuPheILE 335  
|||  
|||  
546 ATGTTGTAATAAGCCTCATGACCTACAACTGGAAGAAATACCCCACTGTGTATTTCATA 605  
QY 336 TyrGIyGIyProGIuValAlaILE 343

Db 606 TATGAGGTCCCGAGTGCACTG 629  
Search completed: October 16, 2003, 03:25:35  
Job time : 2818.69 secs